Treatment targets and outcomes in randomised controlled trials of exercise for non-specific low back pain

Lianne Wood

A thesis submitted for the degree of Doctor of Philosophy

June 2021

Keele University

This thesis is dedicated to my father

Ian James Wicks

11.09.1952 – 5.10.2017

Acknowledgements

This work would not have come to fruition had it not been for the advice, mentorship, guidance and opportunity offered to me by my supervisory team: Professor Nadine Foster and Dr Annette Bishop, in particular, who have mentored me as a physiotherapist and clinical academic, and Dr Martyn Lewis. I am forever grateful and indebted to their expertise and patience with me as I embarked upon the PhD journey, and for their advice and support in developing, undertaking and writing up this thesis.

My thanks also go to all those who made this research possible: the authors who shared their datasets, the participants of the workshops, and the support of my two colleagues (Kieran Bromley and Hollie Birkenshaw) in the management of the final international workshop, as well as the wider workshop team (Jill Hayden, Gert Bronfort and Rob Smeets). Grateful thanks to Dr Jill Hayden, who mentored me through my travel fellowship in Canada, providing invaluable feedback on my systematic review and providing useful discussions around where this research might lead. The opportunity to spend time with your team in Halifax was a pivotal time in my research career. Thanks too, to the Versus Arthritis Primary Care Centre for my PhD fellowship funding, and all at the Versus Arthritis Primary Care Centre at Keele University for their inspiration, support and encouragement. To Ellie Dunstan, thank you for supporting my clinical academic aspirations.

To my parents, for nurturing a love of learning, and for your continued love and support. To my husband Guy, and daughter Mathilde, your support, love and patience have been invaluable through this journey, and your love of adventure inspired me to keep looking forward. I love doing life with you.

Abstract

Persistent non-specific low back pain (NSLBP) is the leading cause of years lived with disability globally, and exercise is the most widely recommended treatment for persistent NSLBP. Previous randomised controlled trials (RCTs) tend to conclude that, on average, exercise has small to medium effects when benefits, versus comparison arms, are judged using the primary outcomes of pain and function. RCTs should select their primary outcome domain(s) and measure(s) based on the rationale of the treatment(s) they are comparing.

This programme of research aimed to i) identify whether existing RCTs match their primary outcome domains to their specified exercise treatment targets, ii) explore whether better matching of primary outcome domains with exercise treatment targets might change the results and conclusions of existing RCT datasets, iii) compare whether composite outcomes composed of multiple matched outcome domains might change the estimates of the between-arm differences through secondary analysis of existing RCT datasets, and iv) gain stakeholder consensus on the treatment targets of exercise interventions in RCTs of persistent NSLBP.

The systematic review included 27 exercise RCTs that, together, stated 31 treatment targets and included six primary outcome domains. Only 25% of included RCTs had primary outcomes that matched their specified treatment targets. Standardised mean differences (SMDs) of exercise versus comparison arms were larger in the matched (SMD 0.54 (95% CI 0.23 to 0.85), p=0.0006) compared to the unmatched category (SMD 0.22 (95% CI 0.01, 0.44) p=0.04), but this difference was not statistically significant (p=0.10).

Abstract

Secondary analyses were conducted on a total of nine previous RCT datasets. First, matching was investigated in five RCTs (n=1033) that used an unmatched primary outcome but included some of their matched outcomes as secondary outcomes, and second, by developing composite outcomes in four RCTs (n=864). Firstly, of five RCTs, three had greater SMDs and increased between-arm statistical significance with matched outcomes compared to an unmatched primary outcome. Of four composite outcomes: three RCTs had greater SMDs and improved statistical precision using the composite outcome compared to the primary outcome in favour of exercise.

Finally, a total of 39 participants contributed to two sequential nominal group consensus workshops. The final prioritised targets of exercise were: improving function, improving quality of life, reducing pain, targeting patient-specific goals, reducing fear of movement and increasing physical activity.

This programme of research has highlighted the need for improved identification and specification of treatment targets of exercise interventions for persistent NSLBP. Matching the primary outcome to the treatment targets of the exercise intervention appears to be important, but composite matched outcomes may be more responsive and require further exploration in RCTs of exercise for persistent NSLBP.

List of Abbreviations

ANOVA	Analysis of variance
ANCOVA	Analysis of variance with covariates
Appts	Appointments
BI	Brief intervention
C	Control
CBT	Cognitive behavioural therapy
CERT	Consensus on exercise reporting template
CG	Control group
CI	Confidence interval
COMET	Core outcomes measures in effectiveness trials
CONSORT	Consolidated standards of reporting trials
CVD	Cardiovascular disease
EQ-5D	Euroqol-5D
Ex.	Exercise
FABQ	Fear-avoidance beliefs questionnaire
FU	Follow-up
GP	General practice/ practitioner
GPE	Global perceived effect
HADS	Hospital anxiety and depression score
HCP	Health-care professional
HEA	Home exercise and advice
HEP	Home exercise programme
HRQoL	Health-related quality of life
IMMPACT	Initiative on methods, measurement and pain assessment in clinical trials
IV	Inverse variance
LBP	Low back pain
MANCOVA	Multivariate analysis of covariance
McK	McKenzie exercise
MCE	Motor control exercise
MCID	Minimum clinically important difference
Min	Minute
MSK	Musculoskeletal
NRS	Numeric rating scale
NS	Non-significant
NSAIDs	Non-steroidal anti-inflammatory drugs
NSLBP	Non-specific low back pain
OA	Osteoarthritis
ODI	Oswestry disability index
OMERACT	Outcome measures in rheumatoid arthritis clinical trials
PDI	Pain disability index
PE	Physical exercise
PMD	Pooled mean difference

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
PSEQ	Pain self-efficacy questionnaire
PSFS	Patient self-functional scale
Pts	Participants
PT	Physiotherapy
Q	Questionnaire
QBPDS	Quebec back pain and disability scale
RCT	Randomised controlled trials
RMDQ	Roland and Morris disability questionnaire
ROM	Range of movement
RTW	Return to work
SD	Standard deviation
SE	Standard error
SET	Specific exercise therapy
SF	Short-form Health Survey
SFA	Solution finding approach
SMD	Standard mean difference
SMT	Spinal manual therapy
SPSS	Statistical package for social sciences
SWD	Shortwave diathermy
TIDieR	Template for intervention description and replication
TSK	Tampa scale of kinesiophobia
UC	Usual care
UCL	Utrecht Coping List
UK	United Kingdom
USA	United States of America
US	Ultrasound
VAS	Visual analogue scale
Wk(s)	Week(s)
WL	Wait –list
WMD	Weighted mean difference
WU	Warm-up
Yrs	Years

I. Table of Contents

Table of Contents

Table of Contents

Table of Contents

List of Tables Included in the Text

List of Tables

List of Tables

List of Tables

List of Tables

List of Figures Included in the Text

List of Figures

List of Figures

List of Equations

List of Equations

1 Background and Introduction

Summary

Persistent non-specific low back pain (NSLBP) is the leading cause of disability worldwide. Exercise is the most commonly recommended treatment for persistent NSLBP, but the treatment targets of exercise are poorly described in the literature. Few randomised controlled trials (RCTs) identify their treatment targets in the justification for their exercise intervention(s). Consequently, they often use primary outcomes that do not reflect the target(s) of the exercise intervention(s). The aim of this thesis is to explore the treatment targets and outcomes of exercise in RCTs of persistent NSLBP.

This chapter summarises the clinical challenge of persistent NSLBP, and the use of exercise as a complex intervention in the management thereof; the possible treatment targets of exercise; the use of RCTs to investigate the effects of exercise compared to other interventions; and challenges in the selection of the most appropriate outcome for RCTs of exercise interventions in the field of persistent NSLBP.

1.1 Low back pain

Low back pain (LBP) is one of the most common musculoskeletal conditions worldwide (1), and most people will experience LBP at some point in their lifetime (2). It is a greater contributor to global disability than any other condition (1–3) and is defined as pain in the area below the lower margin of the ribs and above the gluteal folds, with or without leg pain (4). Non-specific low back pain (NSLBP) is the most common type of LBP experienced in the population (believed to

account for approximately 90-95% of all cases seen in primary care), in contrast to serious spinal pathology such as infection, cancer or fracture (estimated to account for less than 1%), or true nerve root compression (5-10%) (5). NSLBP has traditionally been classified according to the duration of symptoms: acute (<6 weeks), sub-acute (6-12 weeks) and chronic or persistent (>12 weeks). United Kingdom (UK) guidelines (3) recommend a move away from a time-based classification in favour of a continuum where risk factors for poor prognosis are more critical than the duration of symptoms. For this thesis, the term persistent NSLBP is used, with an awareness that NSLBP is not always time-dependent and may describe both persistent NSLBP beyond the initial acute phase (>12 weeks), as well as fluctuating or recurrent NSLBP presentations over time.

In the UK, most patients consulting health-care with persistent NSLBP are managed in primary care, specifically in general practice, where NSLBP accounts for more consultations than any other musculoskeletal pain presentation (6). For most, acute episodes of NSLBP generally improve within the first six weeks regardless of the treatment provided, but some people will experience NSLBP that persists over time (7). Variables influencing the progression to persistent NSLBP include a complex interaction of physical, psychological and social factors (3).

Several clinical guidelines have been produced worldwide regarding the diagnosis, assessment and management of persistent NSLBP (3,8,9). They provide similar recommendations for the management of acute, subacute and persistent NSLBP (10,11). They frequently recommend the use of simple

strategies such as education, reassurance, and safe medications, as well as more complex interventions, such as exercise, multidisciplinary rehabilitation, and cognitive-behavioural therapy interventions (11). Clinical guidelines are tools which provide recommendations for the management of conditions, such as persistent NSLBP. Guidelines are informed by the interpretation of results of RCTs and systematic reviews of RCTs, which provide evidence for the effectiveness of one intervention in comparison to another (or a control arm) (12). However, synthesis of the evidence from RCTs of complex interventions is challenged by the heterogeneous nature of complex interventions, the different settings in which the intervention is delivered, the nonlinear pathways between the intervention and identified outcomes, and the different intervention components (13).

1.2 Exercise as a Complex Intervention

Complex interventions are created when numerous components interact independently and interdependently (14). These components may be difficult to define precisely, and therefore it is more challenging to evaluate what the interactions and relationships are (15). Intervention programme theory (for example, using a logic model) helps to make sense of complex interventions by describing visually: a) the intervention components and the relationships between them, b) the underlying theories of change and causal assumptions between the intervention components and resultant outcomes, and c) the interactions between the intervention and the contextual implementation setting (13). Logic models are increasingly recommended in the evaluation and

assessment of complex interventions, such as exercise, to ensure methodological rigour and facilitate high fidelity implementation (16,17).

Exercise is a sub-group of physical activity, which is planned, repeated, structured and intends to improve one's physical fitness (18). Therapeutic exercise is defined as "the use of active or assisted exercises aimed at improving range of motion, strength, or dynamic neuromuscular control of joint motion"(19). Therapeutic exercise is an example of a complex intervention due to the numerous components that may affect the individual's biological (20), psychological and social (21) functioning, as well as the addition of treatment interaction between exercise providers/ therapists and patients (22). Within exercise interventions, consideration can be given to: the wide variation in the type of therapeutic exercise delivered (for example yoga, motor control retraining, general conditioning exercise), as well as the setting in which it is delivered and undertaken (at a health centre, in a private clinic room, in a community centre, at home etc.), the deliverer of the intervention (e.g. a fitness instructor, physiotherapist, sports trainer etc.), the style of intervention received (group, individual, video format, via the internet), the addition of a home-based programme of exercises, as well as the frequency of the exercise(s) and other components of exercise dose (23). All of these contribute to the complexity in the tasks of synthesising best available evidence about exercise for NSLBP and comparing exercise interventions between different RCTs and clinical settings. Guidelines for persistent NSLBP and sub-acute NSLBP all recommend exercise; however, the advice provided is inconsistent about the type of exercise and mode of delivery or the most effective frequency, intensity or dosage (11). Recent

clinical guidelines recommend a variety of types of exercise for the management of patients with persistent NSLBP, such as tai-chi, sports rehabilitation, strength training, aerobic conditioning, yoga, and motor control exercise (11). However, patient preference and therapist recommendation are also considered to influence the exercise programme prescribed.

1.3 Exercise

Regular and adequate levels of physical activity have significant health benefits at all ages: improving muscular and cardiorespiratory fitness, bone and functional health; preventing falls and depression, improving general cardiovascular health; and maintaining energy levels and weight control (18). Exercise appears to reduce the severity of persistent pain as well as to act more generally, leading to improved overall physical and mental health, and physical functioning (21).

In the management of persistent NSLBP, all exercise types have been shown to have, on average, some benefit on pain and function, with few adverse effects and low associated costs (3). Stabilising and resistance exercise appears to result in the greatest improvement in physical function (24). Strengthening and stretching programmes demonstrate the greatest reductions in pain intensity compared with other exercise types, in the most recent Cochrane review of exercise for LBP (25), while Pilates, aerobic exercise and motor control exercise appeared to have the best effect on pain intensity in a more recent network meta-analysis (24). Individually designed exercise programmes, supervised home exercise programmes and individually supervised programmes, compared with

home exercises alone, appear to have a more favourable effect on pain scores in persistent NSLBP (25). Furthermore, high dosage exercise programmes (more than 20 hours' intervention time) appear to be more effective than low dosage programmes and interventions including other types of conservative care (25). Exercise therapy appears to positively influence fear-avoidance beliefs, pain catastrophising beliefs and self-efficacy expectations (22).

Whilst the above highlights that exercise is beneficial for NSLBP, in general best evidence highlights that the size of these benefits are small to modest. Exercise for persistent NSLBP yields small benefits when compared to all controls, with a pooled mean improvement of 7.3 points (95% CI, 3.7 to 10.9) for pain (out of 100) and 2.5 points (1.0 to 3.9) for function (out of 100) at the earliest follow-up time-point (26). More recently, a systematic review and meta-analysis (of 45 RCTs) found a standardised mean difference (SMD) of -0.32 (95% CI -0.44, -0.19) in favour of exercise in comparison to all controls (27) (medium effect size (28)). Strong evidence has also been found in favour of exercise for NSLBP with small to medium effects on pain, physical function and quality of life in comparison to controls (29). However, the results of these recent systematic reviews show some inconsistency, and the evidence base considered as low quality due to the inclusion of small sample sizes and therefore, potentially underpowered studies (21,26). So overall, despite strong evidence (i.e. many RCTs testing exercise), it appears that, at best, exercise leads to small to medium average benefits compared to alternative treatments or no treatment.

1.3.1 Treatment Targets and the Mechanism of Action of Exercise

Although exercise is widely recommended, the treatment targets and mechanisms of action of exercise in persistent NSLBP are not fully understood. Exercise may work through many potential mechanisms of action, including some or all of the following: peripheral effects (such as targeting musculoskeletal impairments such as weakness in the muscles in the back or lack of flexibility of the spine), central effects (such as through the neural systems), psychological effects (such as addressing fear-avoidance beliefs or self-efficacy expectations) or social effects (such as reducing social isolation). The mechanisms underpinning the persistence of NSLBP are similarly poorly understood, despite a variety of models which have attempted to explain its persistence. Understanding the process whereby persistent NSLBP is maintained may help to identify important targets for interventions. Verbunt et al. (30) describe two different models that may explain contributing mechanisms to persistent NSLBP via the 'disuse syndrome'(31): the fear-avoidance model[1] (32) and the suppressive model[2] (33). These both highlight potential mechanisms which may sustain persistent NSLBP when disuse is present – including physical, psychological and social changes resulting in deconditioning.

[1] The fear-avoidance model was first described by Lethem et al (335) and helped to explain the persistence of pain in the absence of pathology whereby individuals have an increasing disuse/disability/depression related to avoidance of the aggravating activity initiated by a fear of pain.

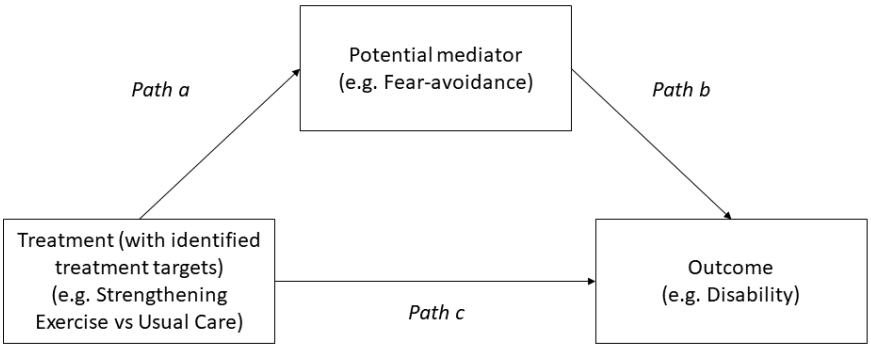
[2] The suppressive model describes a sub-group of individuals who in addition to the avoidance strategies, cope with pain using endurance strategies: they ignore the pain and then overload their muscles which leads to muscular hyperactivity (30).

Systematic reviews confirm the lack of strong evidence about the relationships between specific musculoskeletal impairments (physical changes) and pain and disability in persistent NSLBP (22,34,35). It does not appear that strengthening, motor control or flexibility improvements in specific ranges and muscles are correlated with outcomes, such as pain and physical function (22,34,35). Alternative mechanisms for exercise may include neurobiological mechanisms, such as through gene expression (36,37), endogenous opioid pathways, nociceptive inhibitory pathways (38), anti-inflammatory pathways (39) and altered hormone levels (36,40).

Other explanations suggest exercise may work through central effects, through correcting an altered body schema, reweighting sensory input which may modify motor control patterns, or through a positive patient-therapist relationship (22,41,42). Wand et al. (42) suggest that these features may be epiphenomena of cortical dysfunction instead of an underlying (spinal) mechanism of persistent NSLBP. Exercise may alternatively act through psychological mechanisms, through reducing catastrophising, fear-avoidance beliefs or improving self-efficacy expectations in persistent NSLBP (22,41,42). These different potential mechanisms suggest that the specific type of exercise used might be largely irrelevant in achieving positive benefits from exercise (22), a suggestion that would fit with the best available evidence and clinical guideline recommendations (3,9). Still, there is a lack of research that fully explains the mechanisms of action and treatment targets of exercise for persistent NSLBP.

A treatment target is defined as the aim of the intervention or treatment (43). Treatment targets should reflect the active ingredients of the intervention, as explained by the mechanisms of action. In certain situations, treatment targets may be the same as or similar to surrogate markers, or intermediate variables, whereas, in others, they may have more direct effects on the intended 'end' outcomes. These terms are conceptually different, yet they are terms that are sometimes used interchangeably. For this thesis, the term treatment targets is used, with the acknowledgement that it is unknown whether treatment targets have an indirect effect on the causal pathway, similar to the action of a mediator (44,45) as demonstrated in Figure 1-1.

Figure 1-1: Example of a mediation model with treatment targets (adapted from Mansell et al. 2013)



Direct effect of treatment on outcome = path c

Mediating effect = ab (combination of paths a and b)

Total effect = ab + c

Despite considerable evidence that exercise is beneficial for patients with persistent NSLBP, there is still uncertainty regarding the treatment targets of exercise for this population (46–48). Rainville et al. (49) proposed that the clinical use of exercise for LBP could be defined in three goals: improving function (through affecting strength, flexibility and cardiovascular endurance), reducing

back pain and decreasing disability through reducing excessive fears related to movement. However, this proposal does not appear to have included any stakeholder or patient involvement and is not based on the mechanisms literature. Helmhout et al. (48) expanded on the goals suggested by Rainville et al. (49) by suggesting three different domains of potential exercise treatment targets: i) mechanical (such as spinal strengthening or stabilisation-based theories), ii) neurological (such as desensitisation theories), and iii) psychological or cognitive theories involving cognitive and or operant conditioning theories (such as graded activity). These theoretical frameworks are important in both clinical settings as well as research of exercise in NSLBP.

In summary, evidence to date highlights the need for studies to ascertain what the active ingredients (and treatment targets) of exercise for persistent NSLBP are (47,50,51). Further, it is essential that clear mechanisms of action (with identified treatment targets) for exercise interventions in persistent NSLBP are developed with a variety of stakeholder voices (patients, clinicians and researchers) so that an understanding of how the exercise intervention's impact through cause, mechanisms and outcomes configurations is gained (15,52).

1.4 RCTs of Exercise Interventions

Systematic reviews of RCTs of exercise for persistent NSLBP consistently demonstrate small to medium SMDs (26,27,29). RCTs with higher internal validity are more likely to demonstrate smaller SMDs, although 95% of LBP RCTs have been suggested to be underpowered to detect small to medium SMDs (53).

The RCT remains the best research design to compare the effects of interventions by reducing the systematic differences between two or more intervention groups, minimising bias and allowing the ability to discern true difference over that related to chance using statistical means (54,55). The evaluation, interpretation and reproduction of exercise interventions (as a complex intervention) in pragmatic RCTs are challenging due to the numerous components that are involved (55,56): especially when information regarding processes and targets of the interventions are not fully described (14). RCTs of poor methodological quality more frequently demonstrate greater treatment effects (57–60), and LBP RCTs with lower risks of bias have been reported to have SMDs up to 0.20 lower than trials with higher risks of bias (61). Errors in the design of RCTs can compromise or invalidate the conclusion and results of the RCT (62).

In an attempt to improve the quality of reporting of RCTs, the Consolidated Standards of Reporting Trials (CONSORT) statement (63) checklists have been modified to "improve the completeness of reporting and ultimately the replicability of interventions" (63,64). The Template for Intervention Description and Replication (TIDieR) checklist was developed specifically for physical therapy intervention studies to improve the reporting of interventions in RCTs (64). A further exercise-specific reporting checklist, the Consensus on Exercise Reporting Template (CERT) (65) aims to increase the reproducibility of exercise interventions in RCTs. These checklists aim to improve the reporting both of outcomes and interventions in RCTs to facilitate reproducibility and clinical applicability. The CONSORT statement has a category for the explanation of the

proposed rationale for the trial (category 2a); however, there is no mention of whether this rationale should include a clear match between the treatment targets of the intervention(s) and the outcomes utilised (category 6a) (63). Similarly, the CERT does not include consideration of the aim or objective of the exercise intervention, yet this is an important omission as it influences both the design of the exercise intervention and the outcomes used to assess its effectiveness (66). A recent review by Wood, Ogilvie and Hayden (67) has demonstrated that a high proportion of previous RCTs failed to clearly state their exercise treatment aims (64%). Therefore, despite the creation of useful checklists to improve RCT and intervention reporting (e.g. CONSORT, TIDieR, CERT), poor reporting of exercise interventions, their rationale, their treatment targets and how these targets shape decisions about the selection of outcomes, persist (68).

1.5 Outcome Domains and Measures in RCTs in NSLBP

An outcome (or outcome domain) is the construct of interest to be measured, that can be represented by a latent variable such as pain, or physical function, which may not be directly observable (69). An outcome measurement instrument (or outcome measure) is the means used to quantify the construct (69). There are a wide variety of outcome measures and domains used to compare treatments in trials in persistent NSLBP: one review identified over 75 different outcome measures across six different domains (68). In RCTs of LBP, the most frequently reported outcome domains are pain, followed by physical function (70). Whereas in trials of rehabilitation in persistent NSLBP, a higher proportion of physical function domains are reported (68).

Although overall reporting of outcomes of RCTs has improved with time, the incompleteness of reporting outcome measures may introduce performance or detection bias. Selective outcome reporting bias may also be present when authors only report outcomes with favourable result (71,72). Selection bias may subsequently bias results of further systematic reviews or interpretation of RCT results (73). Hence, a recommended standardised set of outcomes aims to reduce performance, selection or detection bias, and therefore improve the quality of trials, both for the benefit of those interpreting the results and for use in systematic reviews and meta-analysis.

Despite recommendations to standardise outcome measures in RCTs of interventions for persistent NSLBP since 1998 (74–78), systematic reviews have demonstrated these have had a limited or inconsistent effect on research practice (70). The initiative of the Core Outcomes Measures in Effectiveness Trials (COMET) provides methodological research and guidance for the development of core outcome sets (79). Core outcome sets refer to the minimum outcomes that should be collected within RCTs in a certain discipline, thus aiding between trial-comparisons, contrasts and meta-analyses (80). Further to the progress of the "Outcome Measures in Rheumatoid Arthritis Clinical Trials" (OMERACT) group (81), COMET influenced the most recent work on the development of the core outcome set for LBP (78) The "Initiative on Methods, Measurement and Pain Assessment in Clinical Trials" (IMMPACT) group recommended six core domains when designing clinical trials testing treatments for chronic pain: pain, physical function, emotional function, participant ratings of improvement and satisfaction, symptoms and adverse events, and participant disposition (76). More recently,

Chiarotto et al. (78) gained consensus on pain intensity, physical function and health-related quality of life (HRQoL) amongst health-care providers, health-care researchers and patients as core outcome domains for RCTs comparing interventions for persistent NSLBP. Further consensus work by Chiarotto, Terwee and Ostelo (82), recommended outcome measures for each agreed outcome domain, including the Oswestry Disability Index (ODI) (version 2.1) for physical function, the Numeric Rating Scale (NRS) for pain, and the Short-Form (SF) 36 for HRQoL. It is important to note that these recommendations for core outcome domains and measures for RCTs in the field of LBP do not stipulate which outcome(s) should be used as the primary outcome for an RCT, nor do they exclude the additional use of other outcome measures in an RCT (71).

The challenging choice of outcome measure selection is encapsulated by the following quote:

"The ultimate value of a RCT …will be directly tied to how well the selected outcome measure matches the researcher's understanding of what he or she expects to change, to what degree it is expected to change, over what period of time this change will happen and how that change can best be identified" (83).

Considering exercise as a complex intervention with multiple potential treatment targets, it follows that multiple outcomes or an outcome that is multi-domain in nature might be more appropriate for RCTs testing exercise. However, the use of multiple outcomes in RCTs has been cautioned against (84). The concerns expressed include the situation where the intervention yields statistically significant benefits over the control on one of the selected outcome measures and thus a conclusion of superiority is drawn, even though the intervention was

not superior when assessed using the other outcome measures (84). RCT analysis plans should test a working hypothesis based on the treatment targets of the intervention, which should lead to the selection of the primary outcome measure, from which the minimally important difference can be used to calculate the sample size (84). Most literature regarding trial design stipulates that the primary outcome should match the rationale of the intervention (52,71). However, the selection of the primary outcome in a trial may be made based on a variety of factors, with increasing pressure to choose outcomes in line with patient preference and core outcome domains. Further, in complex interventions such as exercise, which frequently have more than one treatment target, the selection of one single primary outcome measure may be insufficient to reflect the wide range of treatment targets required to investigate superiority (85). A potential reason for the small between-arm effect sizes seen in RCTs of exercise interventions in persistent NSLBP is the poor matching of RCT outcome domains with the exercise treatment targets, highlighted in the past by Chapman et al. (86), Gilron and Jensen (87) and Helmhout et al. (48).

A combination of outcome measures may provide a better examination of the benefits of exercise versus control for persistent NSLBP, rather than focusing only on one primary outcome measure. This has been the approach in the field of osteoarthritis (81), as well as in rheumatoid arthritis (88,89) and diabetes (90). The OMERACT group (81), for example, reached consensus that in order to conclude ‘treatment success' for arthritis, an improvement in pain, physical function and global improvement were all required (75). The use of a combination of outcome measures may be more clinically meaningful than reporting only the

estimate of between-arm difference based on a single outcome measure; however, due to the limited evidence on composite measures available for LBP future research has been recommended (75). In persistent NSLBP, a composite index of response has been proposed and includes items in the domains of pain, function and overall impression of health (91). However, this index does not appear to be widely used. Ultimately, further studies are required to refine a responder index and or advise on composite outcomes in persistent NSLBP, potentially incorporating the core outcome sets defined by Chiarotto et al. (82).

1.6 Rationale for PhD

The previous sections in this chapter have highlighted the gaps in knowledge that this thesis has been designed to address. In summary, exercise is a commonly recommended 'core' treatment for patients with persistent NSLBP, based on the best available evidence from systematic reviews of RCTs. While exercise is an excellent example of a complex intervention with multiple treatment targets, most RCTs focus primarily on a limited set of primary outcomes driven by international consensus, namely reduction in pain and increase in function. The focus on pain and physical function as the key primary outcome domains in RCTs may not reflect the breadth of benefits that exercise can provide. This focus has resulted in a mismatch: the most commonly used outcome domains in RCTs comparing treatments for persistent NSLBP do not match the breadth of treatment targets of exercise interventions.

A clearer understanding of the key treatment targets of exercise interventions for patients with persistent NSLBP is required, in line with recommendations for further research (92). Research that explores the importance of matching exercise treatment targets and the key outcome domains in RCTs is needed as is research that tests whether better matching of these might change the results and conclusions of exercise RCTs (48,84). The next chapter describes the aims and objectives of this PhD programme of research.

2 Chapter 2: Aims, Objectives and Research Design

Summary

This chapter provides an overview of the thesis and chapter aims and objectives, as well as providing an overview of publications and presentations from, and relating to, this thesis.

2.1 Introduction

This chapter presents the overall aim of the thesis. Included is a brief synopsis of each of the subsequent chapters and a schematic overview of the thesis. Four stages of research were planned to address the overall aim: each requiring different research designs, methods and analyses. The aims, objectives and research design for each of these stages and chapters are described. This is followed by a list of the dissemination activities, to date, resulting from the research described in the thesis and additional publications currently in preparation.

2.2 Overall Aim and Objectives

The overall aim of this doctoral thesis is to investigate whether better matching the outcome domains used in RCTs in the field of persistent NSLBP to the treatment targets of the exercise interventions being investigated might change the results and conclusions of these trials. To achieve this aim, this thesis presents work undertaken in four stages, with the objectives for each stage as follows:

Overall Objectives:

Stage 1: To identify whether existing RCTs of exercise interventions in persistent NSLBP match the primary outcome domains to the identified treatment targets of the exercise intervention by systematically reviewing the available literature.

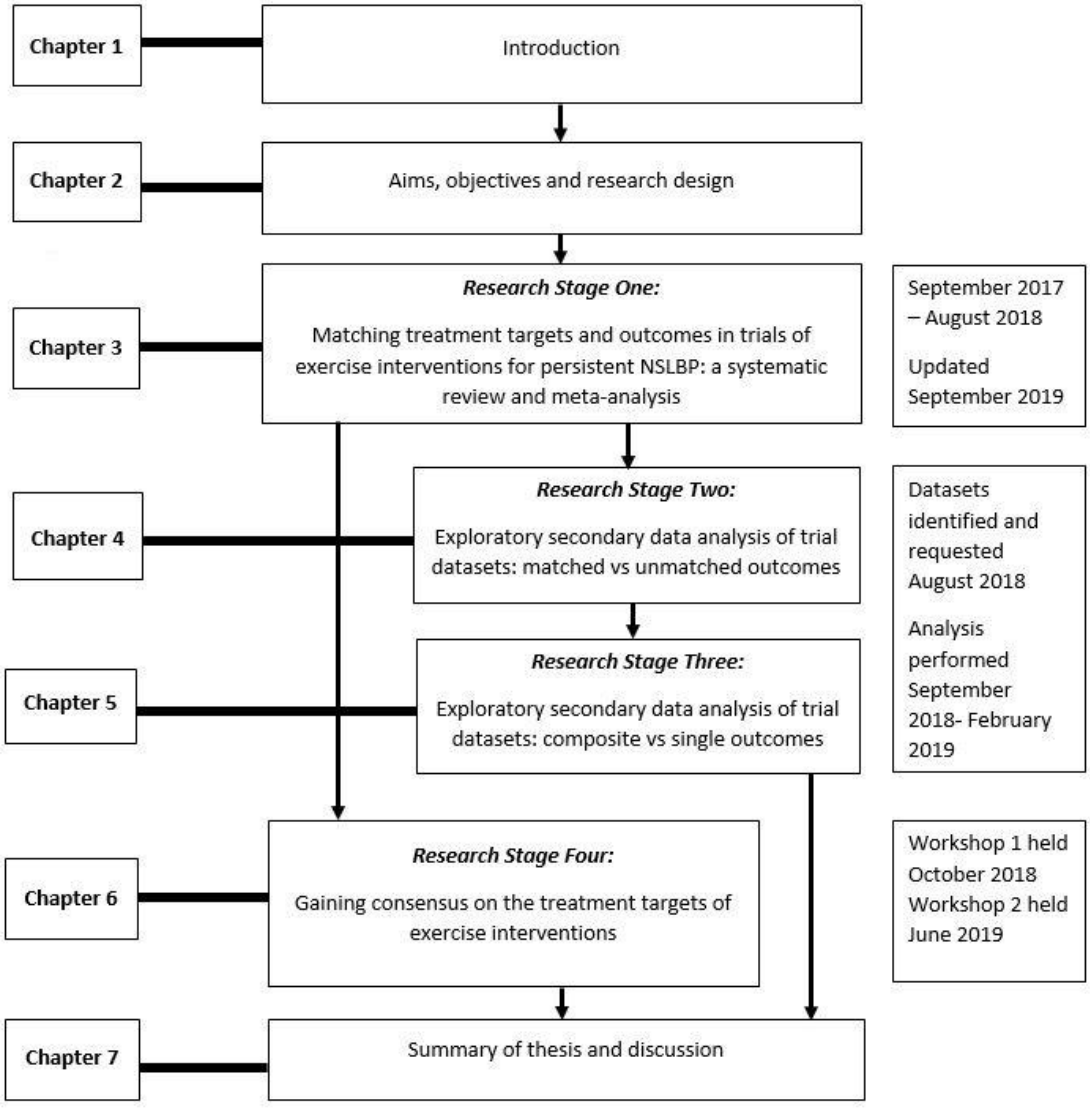
Stage 2: To explore whether better matching of primary outcome domains with exercise treatment targets might change the estimates of the between-arm differences in existing RCTs, and therefore change the results and conclusions of the RCTs, through secondary analysis of existing RCT datasets.

Stage 3: To investigate whether the use of composite outcomes composed of multiple matched outcome domains in comparison to single primary outcome domains might change the estimates of the between-arm differences in RCTs of exercise for persistent NSLBP, through secondary analysis of existing RCT datasets.

Stage 4: To gain stakeholder consensus on the treatment targets and prioritisation of treatment targets of exercise interventions in RCTs of persistent NSLBP through the use of consensus methods.

Each of these stages and the overall structure of this thesis are presented in the schematic overview in Figure 2-1.

Figure 2-1: Schematic overview of the programme of research, timelines and thesis structure



2.3 Summary of Thesis Chapters

The chapters of this thesis are as follows:

Chapter 1 – Introduction

This chapter describes the background and justification for this programme of research.

Chapter 2 – Aims, Objectives and Study Design

This chapter details the overall aim and specific objectives of the programme of research and summarises the content and structure of this thesis.

Chapter 3 - Matching treatment targets and outcomes in trials of exercise interventions for persistent NSLBP: a systematic review and meta-analysis

This chapter describes a systematic review of RCTs of exercise interventions in persistent NSLBP to investigate the role of matching between reported outcome domains and exercise treatment targets.

Specific Objectives:

i. To systematically identify and select RCTs investigating exercise, as an example of a complex intervention, for patients with persistent NSLBP to synthesise the evidence regarding treatment targets, outcome domains and outcome measures.
ii. To assess the risk of bias of included RCTs by utilising the Cochrane Risk of Bias tool (93).
iii. To describe the relationships between reported treatment targets and outcome domains used in these RCTs.
iv. To describe the effect sizes of exercise versus comparison/control treatment for individual RCTs and calculate these when not reported.

v. To define the treatment targets and outcome domains within each RCT, using a process-orientated logic model.

Chapter 4 - Exploratory Secondary Data Analysis of Trial Datasets

This chapter describes secondary analyses of RCT datasets to explore the hypothesis that matching the primary outcomes to the treatment targets of exercise in persistent NSLBP trials may alter the results and conclusions of RCTs.

Specific Objectives:

i. To perform the analysis applied to the primary outcome(s) by the authors of the RCTs on the included secondary outcome(s) that match their stated exercise treatment targets.

ii. To compare the results of the calculated standardised mean differences (SMDs) using matched secondary outcomes with the SMDs of the nominated primary outcomes of RCTs.

Chapter 5 - Exploratory Development of Composite Outcomes in Exercise Trials for NSLBP

This chapter describes secondary analyses performed on datasets to explore the hypothesis that the creation of a composite primary outcome matched to the treatment targets of exercise in persistent NSLBP RCTs alters the results and conclusions of the RCTs.

Specific Objectives:

i. To replicate the analysis applied to the primary outcome(s) by the authors of the RCTs or, where these were not available, on the secondary outcome(s).

ii. To calculate a composite outcome standardised mean difference (SMD) using standardised averages of outcomes matched to the treatment targets of the exercise intervention per identified RCT.

iii. To calculate a co-primary composite outcome SMD using standardised averages of the primary outcomes where nominated by the trials' authors.

iv. To compare composite and co-primary composite outcome SMDs with the reported primary outcome SMDs per identified RCT.

Chapter 6 – Gaining Consensus on the Treatment Targets of Exercise Interventions

This chapter describes the results of two sequential consensus workshops that ranked key treatment targets for exercise interventions in persistent NSLBP in order of priority, identified from the results of the systematic literature review and by prescribers, users of and developers of exercise interventions in the field of persistent NSLBP.

Specific Objective:

To reach consensus on the treatment targets of exercise interventions in persistent NSLBP in order of importance by using nominal group workshops.

Chapter 7 – Summary of Thesis and Discussion

This chapter summarises the key findings of all stages of this PhD programme of research, discusses what these add to knowledge in the field, highlights the strengths and weaknesses of the research, as well as potential implications for future research and clinical practice.

2.4 Methodological Overview of Thesis

This PhD programme of research comprises an exploratory body of work using different research methods: the first three research stages are quantitative in nature, using existing RCT data in the systematic review and secondary data analyses. The final stage of this thesis uses both quantitative and qualitative methods in the form of consensus workshops. All analyses performed within this thesis have been undertaken by the PhD candidate independently, with supervision guidance where needed from my statistical supervisor, Dr Martyn Lewis. The PRISMA guidelines (94) informed the systematic review (stage one).

2.5 Specific Training Undertaken to Complete this PhD

During the course of this PhD, the PhD candidate completed two statistical training modules (MMedSci Statistics and Epidemiology 2017; MMedSci Advanced Quantitative Data Analysis 2018) and a five-day course on Complex Evaluation Development and Evaluation in 2018 (DECIPHer Evaluating Complex Public Health Interventions). Informal training was undertaken through the School for Primary, Community and Social Care on performing systematic reviews (a series of three mini-workshops between December 2017 and March 2018) as well as advanced training workshops on the use of Microsoft Word, Excel and Powerpoint.

2.6 Publications, Presentations and Awards

The following summarises the dissemination activities from this research to date, along with further dissemination plans.

Peer-reviewed journal papers from this thesis

Wood L, Foster NE, Lewis M, Bishop A. Exercise interventions for persistent non-specific low back pain – does matching outcomes to treatment targets make a difference? A systematic review and meta-analysis. Journal of Pain, 2020, Jun 21; S1526-5900(20)30039-0. doi: 10.1016/j.jpain.2020.04.002.

Wood L, Bishop A, Lewis M, Smeets RJEM, Bronfort G, Hayden JA, Foster NE. Achieving consensus on the treatment targets of exercise in persistent non-specific low back pain: a modified nominal group workshop process. Journal of Physiotherapy*, accepted.*

Wood L, Foster NE, Lewis M, Bronfort G, Groessl E, Hewitt C, Miyamoto G, Reme SE, Bishop A. Matching the outcomes to treatment targets of exercise for low back pain: does it make a difference? Results of secondary analyses from individual patient data of randomised controlled trials and pooling of results across trials in comparative meta-analyses. *Under review.*

Published abstracts from this thesis

Wood L, Foster NE, Lewis M, Bishop A. How well do the treatment targets and outcomes match in trials of exercise interventions for chronic low back pain? A

systematic review of the literature. Orthopaedic Proceedings, 101-B, Supp_9. The Society for Back Pain Research 2018 Meeting, Groningen, the Netherlands, 15-16 November 2018.

Wood L, Foster NE, Lewis M, Groessl E, Bronfort G, Hewitt C, Miyamoto GC, Reme SE, Bishop A. Matching the outcomes to treatment targets of exercise in randomised controlled trials in low back pain: does it make a difference? Results of secondary analysis of randomised controlled trials. Orthopaedic Proceedings, 101-B, Supp_10. The Society for Back Pain Research Annual General Meeting, 2019, Sheffield, United Kingdom, 5-6 September 2019.

Wood L, Foster NE, Lewis M, Bishop A. Achieving consensus on the treatment targets of exercise in persistent non-specific low back pain: a modified nominal group workshop process. Orthopaedic Proceedings, 101-B, Supp_10. The Society for Back Pain Research Annual General Meeting, 2019, Sheffield, United Kingdom, 5-6 September 2019.

Research presented from this thesis

<u>Oral presentations</u>

Wood L, Foster NE, Lewis M, Bishop A. How well do the treatment targets and outcomes match in trials of exercise interventions for chronic low back pain? A systematic review of the literature. Orthopaedic Proceedings, 101-B, Supp_9. The Society for Back Pain Research 2018 Meeting, Groningen, the Netherlands, 15-16 November 2018.

Wood L. Does matching the primary outcome to the treatment goals of exercise treatments change the conclusion of trials in LBP? Institute for Liberal Arts and Science, Keele University, 3 Minute Thesis Presentation. 29th April 2019; Post-graduate Symposium, Research Institute for Primary Care and Health Sciences, Keele University, 16th May 2019.

Wood L, Foster NE, Lewis M, Bishop A. Achieving consensus on the treatment targets of exercise in persistent non-specific low back pain: a modified nominal group workshop process. The International Forum for Back and Neck Pain Research in Primary Care, Quebec City, Canada, 3-6 July 2019; Physiotherapy UK, Birmingham, United Kingdom, 1-2 November 2019.

Poster presentations

Wood L, Foster NE, Lewis M, Bishop A. A systematic review and meta-analysis of the treatment targets and outcomes of exercise as an example of a complex intervention in chronic low back pain. Institute for Liberal Arts and Science, Keele University, 19th April 2018; Post-graduate Symposium, Research Institute for Primary Care and Health Sciences, Keele University, 16th May 2018.

Wood L, Foster NE, Lewis M, Bishop A. How well do the treatment targets and outcomes match in trials of exercise interventions for chronic low back pain? A systematic review of the literature. Orthopaedic Proceedings, 101-B, Supp_9. The Society for Back Pain Research 2018 Meeting, Groningen, the Netherlands, 15-16 November 2018.

Wood L, Foster NE, Lewis M, Groessl E, Bronfort G, Hewitt C, Miyamoto GC, Reme SE, Bishop A. Matching the outcomes to treatment targets of exercise in randomised controlled trials in low back pain: does it make a difference? Results of secondary analysis of randomised controlled trials. Society for Back Pain Research 2019 Meeting, Sheffield, United Kingdom, 5-6 September 2019.

Wood L, Foster NE, Lewis M, Bishop A. Achieving consensus on the treatment targets of exercise in persistent non-specific low back pain: a modified nominal group workshop process. Society for Back Pain Research 2019 Meeting, Sheffield, United Kingdom, 5-6 September 2019; Physiotherapy UK, Birmingham, United Kingdom, 1-2 November 2019; Britspine Conference, Virtual poster presentation, United Kingdom, 10-12 March 2021 (deferred from March 2020 due to COVID-19).

Awards associated with this thesis

Travel Fellowship £3500, Society for Back Pain Research, 2019/20.

This award funded me to travel to Halifax, Nova Scotia, Canada to spend two weeks working with Dr Jill Hayden and her team. The award also funded my travel to Quebec City, Quebec, Canada to host a workshop that contributed to the consensus results (chapter 6), and present the results of our first UK, national workshop. It resulted in a publication and a co-applicant grant award..

Research articles associated with this thesis

Mansell G, Burgess RM, Stynes S, Harrisson SA, Wood L, Hill JC. Investigating change in musculoskeletal patient-reported outcome measures. Journal of Physiotherapy, *under review*.

Wood L, Ogilvie R, Hayden JA. Specifying the treatment targets of exercise interventions in low back pain. British Journal of Sports Medicine, Published Online First: 28th May 2020. http://dx.doi.org/10.1136/bjsports-2020-101981.

Research presented and associated with this thesis

Poster presentation

Wood L, Ogilvie R, Hayden JA. Specifying the treatment targets of exercise interventions in low back pain. Virtual Physiotherapy UK, 14 and 15 November 2020.

2.7 Summary

This chapter presented the overall aim and specific objectives of this PhD programme of research, and the four stages of research that it is comprised of. The next chapter details the aims, methods and results of the first stage of the PhD, the systematic review.

3 Chapter 3: Matching treatment targets and outcomes in trials of exercise interventions for persistent NSLBP: a systematic review and meta-analysis

Summary

This chapter describes the methods, results and conclusions of a systematic review and meta-analysis investigating treatment targets and outcomes in RCTs of exercise interventions for persistent NSLBP. The contents of this chapter have been, in part, accepted for publication as:

Wood L, Foster NE, Lewis M, Bishop A. Exercise interventions for persistent non-specific low back pain – does matching outcomes to treatment targets make a difference? A systematic review and meta-analysis. Journal of Pain, 2020, Jun 21; S1526-5900(20)30039-0. doi: 10.1016/j.jpain.2020.04.002.

3.1 Introduction

As highlighted at the end of the Introduction in chapter 1, the use of a single primary outcome domain and measure in RCTs of exercise for persistent NSLBP may not adequately reflect the multiple treatment targets of this complex intervention. It is unknown whether reports of RCTs of exercise interventions adequately describe the exercise treatment targets, or if these targets are captured by the outcome domains and measures selected. A systematic review of the published literature was conducted to establish what treatment targets,

outcome domains and outcome measures have been used in RCTs of exercise interventions for persistent NSLBP, and to explore whether the primary outcome domains match the treatment targets of the exercise interventions. A systematic search was developed and conducted, RCTs were included based on predefined criteria, and standardised risk of bias assessment and data extraction were completed. A narrative synthesis described the treatment targets, primary and secondary outcomes of included RCTs. The RCTs were classified into sub-groups according to the degree of matching between the primary outcome domain(s) and treatment targets. Standardised mean differences (SMDs) in the primary outcome between intervention groups in each RCT were calculated, and meta-analysis conducted to explore whether matching of outcome domain to treatment targets impact on the size of these SMDs.

3.1.1 Study Aim and Objectives

Aim: The primary aim of this systematic review was to investigate the role of matching between reported outcome domains and exercise treatment targets.

Objectives:

i. To systematically identify and select RCTs investigating exercise, as an example of a complex intervention, for patients with persistent NSLBP to synthesise the evidence regarding treatment targets, outcome domains and outcome measures.
ii. To assess the risk of bias of included RCTs by utilising the Cochrane Risk of Bias tool (93).

iii. To describe the relationships between reported treatment targets and outcome domains used in these RCTs.

iv. To describe the effect sizes of exercise versus comparison/control treatment for individual RCTs and calculate these when not reported.

v. To define the treatment targets and outcome domains within each RCT, using a process-orientated logic model.

3.2 Methods

A protocol of the systematic review was published on the PROSPERO register on the 12th December 2017 (reference CRD42017072023). This review is reported according to the PRISMA statement for systematic reviews (94), ensuring a rigorous approach.

3.2.1 Eligibility Criteria

The following eligibility criteria were applied to identify RCTs for inclusion in the review:

Types of participants: The definition used for persistent NSLBP was symptoms present for more than 12 weeks, with more than 50% of participants having NSLBP. Only RCTs with adult populations were included. The following populations were excluded: those with specific spinal conditions only (such as spinal stenosis, post-surgical pain, pregnancy-related LBP, lumbar instability etc.), serious spinal pathologies (such as cauda equina, spinal tumours, spinal fractures, axial spondyloarthopathy etc.) and widespread chronic pain or systemic pain conditions (such as fibromyalgia and chronic fatigue syndrome).

Types of Interventions: Included exercise interventions were required to be supervised or tailored to fulfil the requirement of a complex intervention. To be eligible, RCTs had to have a comparator group(s) that included no exercise treatment; RCTs may have compared exercise to usual care (e.g. physiotherapy, spinal manipulation, education, or GP-led care), placebo interventions, brief interventions or waitlist controls as long as they did not include a supervised or tailored exercise component. Exercise interventions delivered alongside other active interventions such as manual therapy were excluded as it was not possible to extract the treatment effect of the exercise intervention alone. Cognitive and /or behavioural ('psychological') interventions delivered alongside exercise were included as this component was more difficult to separate from an exercise intervention (23), and exercise for NSLBP is thought to work at least in part through psychological mechanisms of action (see Introduction, section 1.2.1).

Types of comparators: To be eligible, the comparator group(s) were required to have no exercise treatment, and may have received: no health-care, usual care, or be on a waiting list for exercise treatment. Usual care may have included other conservative care such as non-exercise physiotherapy, osteopathic or chiropractic care, manual therapy, drug therapy or other treatments.

Types of outcomes*:* All outcomes relevant to the exercise interventions were included. Outcome domains encompass the focus of the outcome (e.g. 'what' is measured), whereas outcome measures are the measurement instruments for each domain (e.g. 'how' it is measured) (69,78,95). All RCT outcomes (primary and secondary, where stated) were extracted, if they were judged to be potentially relevant to the exercise intervention.

Types of settings: To be included, RCTs had to be conducted in a primary or community care setting, similar to the UK or equivalent in other countries. Secondary care outpatient settings were included, but all in-patient intervention settings (i.e. where patients were required to stay overnight in hospital as part of the exercise intervention) were excluded.

Types of studies: Only RCTs were included in this review. Given the reported effect sizes of exercise from previous systematic reviews, RCTs that lacked sufficient power to detect even a medium effect size between exercise and a comparison group were excluded: for a medium effect size of 0.5 (96) in a two-armed RCT, 120 participants would be required at about 80% power (60 per arm) from an unadjusted analysis. Therefore, 180 participants were required for a three-armed RCT. Three-armed RCTs where the intervention in one arm was not relevant as a comparator arm, or it did not have more than 60 participants, were excluded from the meta-analysis but included in the search results. This crude calculation, of requiring at least 60 participants per arm in RCTs, was used to include trials that were likely to have greater statistical power to detect between-arm differences in their primary outcome measure. These RCTs were also more likely to be of higher quality (53,97).

All titles and abstracts were required to be in the English language to be assessed by the reviewers. Where the full-text was in a language other than English, translation was pursued where possible. Articles were excluded where full-text was not available, or translation was not possible.

3.2.1.1 **Search Strategy**

A computer-aided electronic database search of PsycInfo and CINAHL (EBSCO); Web of Science; AMED, Embase and MEDLINE (Ovid); PEDro and Cochrane Central trials registry was developed in consultation with information specialists and used all keywords and subject headings to explore the most important key concepts: persistent NSLBP, exercise therapy, RCTs. The computer-aided search took place from the conception of the databases until the 18th September 2017. An updated search was performed on the 2nd August 2018 and again on the 5th August 2019. Please see Appendix 9a: Systematic Review Search Terms.

3.2.2 Study Selection

All search results were directly imported into online reference management software (Refworks Proquest), which assisted with deletion of duplicates. A pilot review of both title and abstract hits and full-text screening stages were performed with two reviewers checking five papers each, to pilot test and improve the clarity of the eligibility criteria.

Titles were screened and citations excluded when it was apparent that they did not meet the inclusion criteria. Screening until this point was performed by the PhD candidate (LW). Remaining papers were then exported to the Covidence software package for title and abstract screening. To select full-text papers, two reviewers independently screened the titles and abstracts (LW and AB). When a citation was deemed eligible for inclusion, or a decision was unable to be reached

on the information provided in the title and abstract, the full text was retrieved. Full-text review was completed by three pairs of independent reviewers, with the PhD candidate remaining constant across all pairs - disagreements between reviewers were discussed by the supervisory research team (all reviewers).

3.2.3 Risk of Bias Assessment

The Cochrane Risk of Bias Tool version 1.0 (93) was utilised to assess for risk of bias and includes the following domains: randomisation sequence generation, allocation concealment, blinding of both participants and personnel, and outcome assessors, incomplete outcome data (e.g. drop-outs and withdrawals), and selective outcome reporting. Any additional features felt to be important by the reviewers were included under the domain 'other', such as adherence, compliance, drop-out rates in secondary outcomes. For each domain in the tool, a description of the procedures undertaken in the trial with verbatim quotes was extracted. Independent judgement of the risk of bias for each category was made by two reviewers. In line with instructions for the use of the risk of bias tool, raters categorised each domain of the Risk of Bias tool as 'high risk', 'low risk' or 'unclear risk' if insufficient detail was felt to be present to record a judgement. Risk of bias review was completed by three pairs of independent reviewers, with the PhD candidate remaining constant across all pairs - disagreements between reviewers were discussed by the supervisory research team (all reviewers).

Each item in the risk of bias assessment was considered independently without an attempt to collate or assign an overall score. For incomplete outcome data,

trials were categorised as low risk if they used an intention to treat analysis (63,98). Where this was reported but not performed, trials were categorised as at high risk of bias. Similarly, last observation carried forward, and failure to account for missing data resulted in a high risk of bias classification (99). This was carried out at the same time as data extraction. Graphic representation of the bias judgements was created in an online review manager (RevMan 5.3).

3.2.4 Data Extraction

Pairs of reviewers independently extracted identified treatment targets, primary and secondary outcome domains, and outcome measures from each RCT. Data were extracted utilising a bespoke developed Microsoft Excel Spreadsheet. Further demographic information was also extracted, comprising:

i. The country and setting of the RCT;
ii. Participant characteristics (mean age, proportions of each sex, mean duration of pain);
iii. Total sample size at randomisation, per RCT in total and per arm;
iv. Inclusion/ exclusion criteria;
v. Exercise intervention type (such as general, stretching, strengthening, specific exercises such as the McKenzie method or yoga etc.) and dosage and duration;
vi. Comparator treatment type (usual care, placebo, waitlist control, manual therapy), dosage and duration;
vii. Outcome domains and measures utilised as primary and secondary outcomes;

viii. Follow-up time-points: the specified primary time-point was extracted (100).

3.2.4.1 **Treatment Targets Extracted**

Documented exercise treatment targets (both those described explicitly in the RCT published methods section or protocol paper, and those inferred in the background section of the RCT with quotations where possible). Treatment target extraction was completed by three pairs of independent reviewers, with the PhD candidate remaining constant across all pairs - disagreements between reviewers were discussed by the supervisory research team (all reviewers) (101). Explicit or specific targets were first sought from each RCT paper's methods section under the intervention description. Specific treatment targets were extracted when the authors of the RCT stated that "the exercise aimed to…". If treatment targets were not specified in the methods section, then the background section was scrutinised to search for a specific or implicit description of the exercise intervention targets. If no treatment targets were specified or implied, then related exercise intervention development papers or RCT protocols (where available) were reviewed. Treatment targets were extracted verbatim from the text, and where the identification or specification of treatment targets was unclear or vague, the potential treatment target-related text was extracted.

3.2.4.2 **Outcomes Extracted**

Data extraction focussed on the primary outcomes of each RCT. Outcomes were classified as the primary outcome by using the following process, moving to the next stage if the former did not identify the primary outcome.

i. The primary outcome was explicitly stated by the authors
ii. If more than one primary outcome was used, then the first primary outcome mentioned was used
iii. The outcome measure on which the sample size calculation was based
iv. The first outcome measure referred to in the abstract or paper

This approach has been used in several previous reviews (70,102–105). Secondary outcomes were also extracted when relevant to the proposed treatment targets (e.g. data on adverse events were not extracted). The nominated primary follow-up time-point was also captured, and where this was not specified by the RCT authors, the earliest follow-up time-point post-intervention was extracted and used as the primary time-point (70,104,105). For each primary and secondary outcome, both the outcome domain and outcome measure used were recorded.

3.2.4.3 **Matching of Treatment Targets and Outcomes**

Primary outcome domains were assessed as to whether they matched the RCT authors' own reported exercise treatment targets. The RCT was classified as 'matched' if the primary outcome domain was matched to *any* of the identified treatment targets. RCTs were classified as 'unmatched' if the primary outcome was not matched to *any* of the specified treatment targets. Included in the

unmatched group, were trials which had captured *some* secondary outcomes that reflected the identified treatment targets. These trials were classified in a 'partial matched group' as a sub-group within the unmatched category. This sub-group was useful for identification of trials for future analyses in the thesis, but for the purposes of this chapter are reported as 'unmatched'. This was a subjective process, and was independently judged by two reviewers (of three possible pairs). As this categorisation has not been performed before, there was no formally validated process.

3.2.5 Descriptive Analysis

Following data extraction, data were summarised into tables and analysed. Descriptive statistics were utilised to summarise the number of trials: a) reporting exercise treatment targets, b) measuring outcome domains matched to reported exercise treatment targets and c) measuring outcome measures used within outcome domains. Descriptive statistics, such as frequency counts and percentages, were used to summarise treatment targets, treatment outcome domains and outcome measures.

3.2.6 Meta-analysis of Matched and Unmatched Categories

The primary aim of this systematic review was to investigate the role of matching between reported outcome domains and exercise treatment targets. The meta-analysis was conducted to explore whether matching of outcome domain to treatment targets impact on the size of these SMDs. The analysis was undertaken using the following stages:

i. Individual RCT SMDs for exercise and control arms comparisons

ii. Grouped SMD comparisons

iii. Sensitivity and sub-group analyses

3.2.6.1 **Standardised Mean Difference Calculations**

SMDs were calculated for each trial for between-arm differences at the primary outcome time-point designated by the trial authors, or the earliest time-point post-intervention if no primary time-point was specified. The SMD calculation used this formula:

$$d = \frac{\dot{X}1 - \dot{X}2}{s} \qquad 3\text{-}1$$

Where *d* represents SMD, $\dot{X}$ represents the mean follow-up score, 1 represents the intervention arm, and 2 represents the control arm, and s represents the average of the baseline standard deviations (106). The average was calculated as the sum of the baseline standard deviations divided by 2.

Where a confidence interval (CI) was provided for the individual arm baseline values without a standard deviation (SD) value, the SD for the individual arm was calculated using the following formula (101):

$$s = \frac{\sqrt{n} \times (upper\ limit - lower\ limit)}{3.92} \qquad 3\text{-}2$$

Where S represents the mean standard deviation of the individual arm baseline values and *n* the sample size for the relevant individual arm.

CIs for the SMD were calculated using the following sequence of equations (107):

$$s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \quad \textit{3-3}$$

(derives a pooled variance)

Where s represents the pooled mean standard deviation of each arm, n represents the sample size, and s^2 is the pooled variance.

$$se = s \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad \textit{3-4}$$

(derives a pooled standard error)

The limits of the CI for the between-arm SMD were then calculated:

$$95\%\ Confidence\ Interval = se \times 1.96 \pm d \quad \textit{3-5}$$

Where d is the estimated/ observed between-arm standardised mean difference

The limits of the CIs for the SMD were then extracted using the above formula for SMD derivation.

<u>Where proportions were reported, the following were used:</u>

The formulae below are given for proportions, but presentation of results are given as percentages (through x100% conversion).

$$\text{Proportion} = \frac{proportion\ intervention+proportion\ control}{2} \quad \textit{3-6}$$

$$s = \sqrt{p\ (\frac{1-P}{n})} \quad \textit{3-7}$$

Where p is the average proportion of intervention, control; n is the sample size

$$d = \frac{proportion\ intervention - proportion\ control}{standard\ deviation} \quad 3\text{-}8$$

Where medians were used:

The average lower/upper quartile across the intervention/ control were calculated with the formula:

$$\text{Average quartile} = \quad 3\text{-}9$$

$$\left(\frac{Intervention\ ((Upper\ quartile - median) + (median - lower\ quartile))}{4} + \frac{Control\ ((Upper\ quartile - median) + (median - lower\ quartile))}{4}\right)$$

This was then multiplied by $\frac{34}{25}$ to gain the difference between an interquartile marker and a SD: 1SD covers the middle 68% of the normal distribution, whilst an inter-quartile range (upper quartile minus lower quartile) represents the mid-spread (middle 50%) of the data distribution. The final figure is then comparable with a SD value especially the closer the distribution of scores resembles a normal distribution[3].

For the SMD calculation:

Median values were used here to approximate the mean values (when means are not provided) for purposes of SMD estimation, and will be a good approximation the closer the data is symmetric around the median.

$$d = \frac{median\ intervention\ arm - median\ control\ arm}{standard\ deviation} \quad 3\text{-}10$$

Where *d* represents the SMD

[3] For the purposes of comparing available data this method was used, with the assumption that this data was of a normal distribution even though the likelihood was that it was not normal resulting in the use of medians.

Since SMD statistics for all between-arm differences are given based on intervention arm minus control arm then positive SMDs indicate higher values for the intervention (lower for the control). By contrast, negative SMDs indicate lower values for the intervention (higher for the control). Since the direction of scale data may be conflicting (i.e. higher values indicate worse health outcome status for some scales and better health status for other scales), for the purposes of standardisation and ease of evaluation and interpretation within the meta-analysis all SMDs were scaled such that positive SMDs reflect better outcome for the intervention and negative SMDs reflect worse outcome, by multiplication with minus one where necessary. SMDs were interpreted according to Cohen's (96) recommendations, where an effect size of <0.2 is considered ‘small’, around 0.5 is considered ‘medium’ and >0.8 ‘large’.

Where more than one intervention or comparator arm was used:

If an RCT reported results for two separate intervention arms that had similar treatment targets, or more than one control arms, the results were combined according to the guidance described in the Cochrane Handbook (101) using the formulae in

Table 3-.

Table 3-1: Method for combining arms

	Arm 1	Arm 2	Combined Arms	Equation Number
Sample Size	n_1	n_2	n_1+n_2	*3-11*
Mean	M_1	M_2	$\frac{n_1M_1 + n_2M_2}{n_1 + n_2}$	*3-12*
Standard Deviation	SD_1	SD_2	$\frac{\sqrt{(n_1 - 1){SD_1}^2 + (n_2 - 1){SD_2}^2 + \frac{n_1n_2}{n_1 + n_2}(M_1^2 + M_2^2 - 2M_1M_2)}}{n_1 + n_2 - 1}$	*3-13*

3.2.6.2 **Treatment Success**

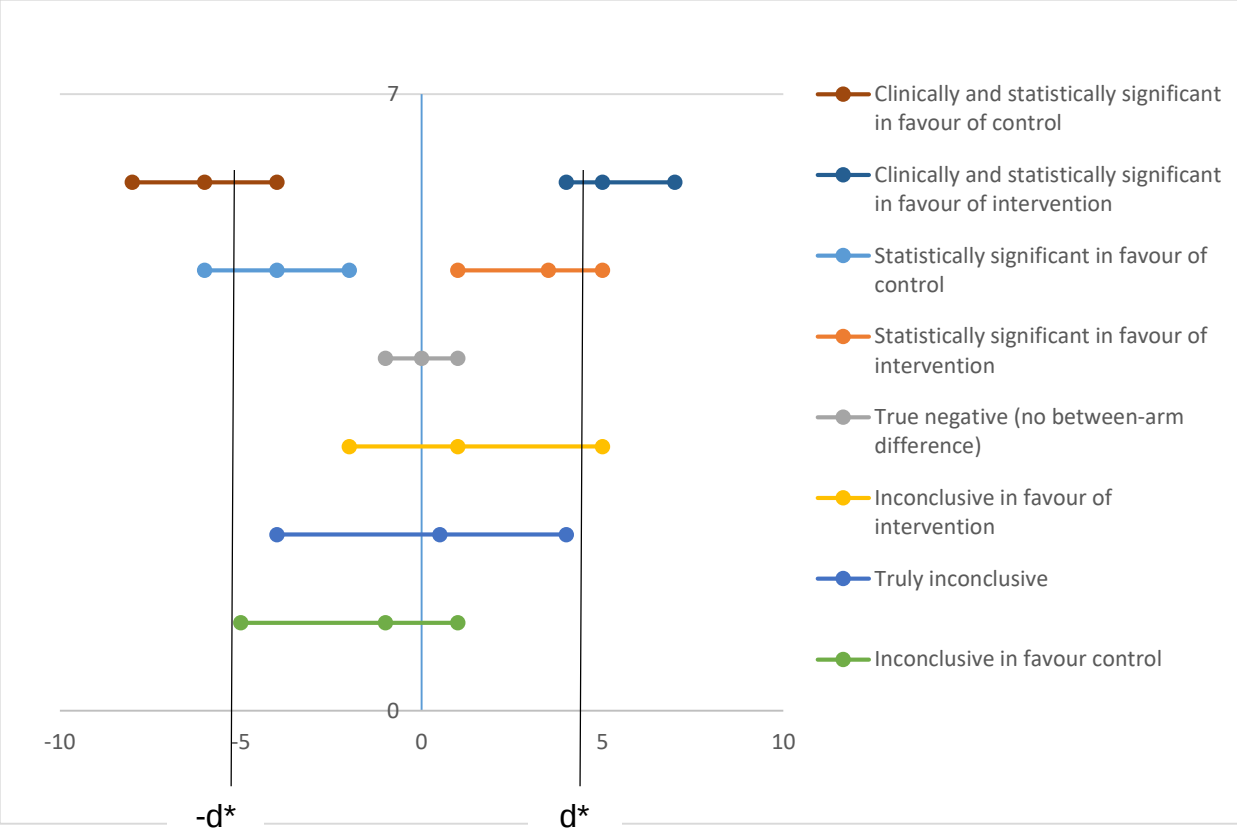
Where possible, SMDs were compared to the reported minimum clinically important difference (MCID), which was standardised by:

$$SMCID = \frac{minimal\ clinically\ important\ difference}{reported\ standard\ deviation\ of\ the\ MCID} \qquad 3\text{-}14$$

SMCID is the standardised minimum clinically important difference, and *MCID* is the minimum clinically important difference.

To determine whether the SMD was clinically meaningful or not, treatment success was evaluated according to the recommendations by Dent and Raftery (108) as demonstrated in Figure 3-1. This figure was amended with the addition of the “clinically and statistically significant” marker which occurs when the SMD and its confidence intervals are greater than the MCID.

Figure 3-1: Modified graphic representation of interpretation of treatment success (adapted from Dent and Raftery, 2011)



*Where d represents the standardised minimum important difference

3.2.6.3 **Grouped Standardised Mean Differences**

Sub-group analysis was performed according to the categorisation of 'matched' and 'unmatched' trials using the RevMan 5.3 software, according to the guidelines in the current version of the Cochrane Handbook for Systematic Reviews of Interventions (101). The Mantel-Haenszel method was used for the random-effects model, where statistical heterogeneity was observed (I^2>=50% or P<0.1). The I^2 defines the amount of heterogeneity across trials in the meta-analysis and is categorised as low (25% or less), moderate (50%) or high (75%

or more) (109). Sub-group analyses were performed to explore the effect on the size of the SMD depending on whether the RCT was matched or unmatched for:

i. The impact of risk of bias assessment
ii. The difference in recruitment strategy (e.g. consulters vs non-consulters)
iii. The impact of specified exercise treatment targets
iv. The effect of the comparator group.

A prediction interval for the estimated heterogeneity levels was calculated as recommended by the Cochrane Handbook (106) using the calculation:

$$95\%\ Prediction\ Interval = \mathrm{M} \pm t_{k-2}\ \times \sqrt{Tau^2 + \mathrm{SE}(M)^2} \qquad 3\text{-}15$$

Where M is the summary mean from the random-effects meta-analysis, t_{k-2} is the 95% percentile of a t-distribution with k–2 degrees of freedom, k is the number of studies, Tau^2 is the estimated amount of heterogeneity and SE(M) is the standard error of the summary mean.

3.2.6.4 **Sensitivity Analyses**

3.2.6.4.1 ***Sensitivity Analysis: Ratio of Means***

A ratio of means (RoM) value was calculated for the primary outcome of each RCT. SMDs were recommended for standardising scales when comparing across different outcome scales which measure the same construct (e.g. combining the McGill Pain score with the Brief Pain Inventory and the Visual Analogue Scale (VAS)) (106). The use of RoM is proposed as an alternative to using the SMD as it can combine outcomes measured using different scales and domains (110). The RoM expresses the percentage change in the mean value of the intervention arm in comparison to the control arm. It has been suggested to be a more favourable method as it is easier to interpret and has

low levels of bias (111). The results are interpreted, similar to the risk ratio, where if the combined ratio is 1.15, then the outcome in the intervention arm is 15% higher than the control arm. The RoM was calculated by dividing the mean outcome of the intervention arm by the mean outcome of the control arm as seen in the equation below.

$$\log(RoM) = log\frac{X_T}{X_C} \qquad 3\text{-}16$$

Where RoM represents ratio of means, X_T is the mean of the intervention arm, X_c is the mean of the control arm.

For meta-analysis, the natural logarithm of each trial RoM and standard error (SE) were calculated using the mean values, number of participants (*n*) and SD in each arm as:

$$SElog(RoM) = \sqrt{\frac{1}{n_T}(\frac{s_T}{X_T})^2 + \frac{1}{n_T}(\frac{s_C}{X_C})^2} \qquad 3\text{-}17$$

Where SE represents standard error, n is the sample size, *s* is the standard deviation and X is the mean of the arm of interest (T is the intervention arm; C is the control arm).

The natural logarithm transformed ratios were then combined across RCTs using the standard inverse variance method. In the inverse variance (IV) method the weight given to each study is the inverse of the variance of the SMD (i.e. the reciprocal (one over) the square of its SE). Hence, within the aggregated ratio, the larger studies have more weight than smaller studies (as the latter have larger SEs). However, in random effects meta-analysis, as heterogeneity increases, the IV weights become more equal, therefore increasing the estimator variance (112). A combined ratio and its 95%

confidence interval were obtained by back-transforming the combined log-transformed ratio and its 95% CI:

$$\mathrm{RoM} = \exp(\log(RoM)_{pooled}) \quad \textit{3-18}$$

$$95\%\ CI = \exp\log(RoM)_{pooled} \pm 1.96 \times SE(log(RoM)_{pooled}) \quad \textit{3-19}$$

These values were calculated, input into RevMan (5.0) and reviewed in comparison to the original SMD values. They were interpreted similarly to SMD values with 8, 22 and 37 percentage values corresponding to Cohen's (96) SMD values of 0.2, 0.5 and 0.8 (110).

3.2.6.4.2 *Sensitivity Analyses: Pain and Physical Function Scores*

Sensitivity analyses were performed by converting reported pain and physical function scores to mean differences on a 0-10 scale for pain and 0-100 scale for function. These were interpreted according to MCIDs within sub-groups using the recommendations of Ostelo et al. (113).

Pain scores were converted to a 0-10 scale, whereby Visual Analogue (VAS) or Numeric Rating Scale (NRS) scores from 0-100 were divided by 10 to translate to a 0-10 scale. Ostelo et al. (113) report that 2 points represent the MCID for the NRS on a 0-10 scale.

Physical function scales were all converted to a 0-100-point scale by multiplying the reported mean score by 100 and dividing by the available scale points (e.g. the Roland Morris Disability Questionnaire (RMDQ) ranges from 0 to 24 thus scores were divided by 24). Ostelo et al. (113) report an MCID of 10 points for

the Oswestry Disability Index (ODI) (0-100), and 20 points for the Quebec Back Pain and Disability Scale (QBPDS) (0-100). As all scales were converted to a 100-point scale, 10 points was the lowest of the above MCIDs and were used as a marker of MCID.

3.2.7 Logic Model

Logic models have been recommended in systematic reviews of complex interventions to allow visual depictions of the interacting components (13,114). A preliminary draft of a logic model was used to assist with interpretation and understanding of the relationships between the treatment targets of exercise and the outcomes used in trials of exercise interventions. A logic model for exercise as an example of a complex intervention in persistent NSLBP was constructed underpinned by the extracted data (13) of the included RCTs. This included the setting and deliverers of the exercise intervention, the treatment targets identified, the core components of the exercise interventions, and the primary outcome domains used to assess the effectiveness of the exercise interventions (13,114,115) as extracted by the PhD candidate. Relationships between reported treatment targets and treatment outcome domains were mapped and presented to the supervision team for refinement.

3.2.8 Publication Bias

Typically, a funnel plot would be created to test for publication bias, but given that small RCTs of exercise for NSLBP were purposely excluded from this review, this would be hard to interpret, and thus this was not created.

3.3 Results

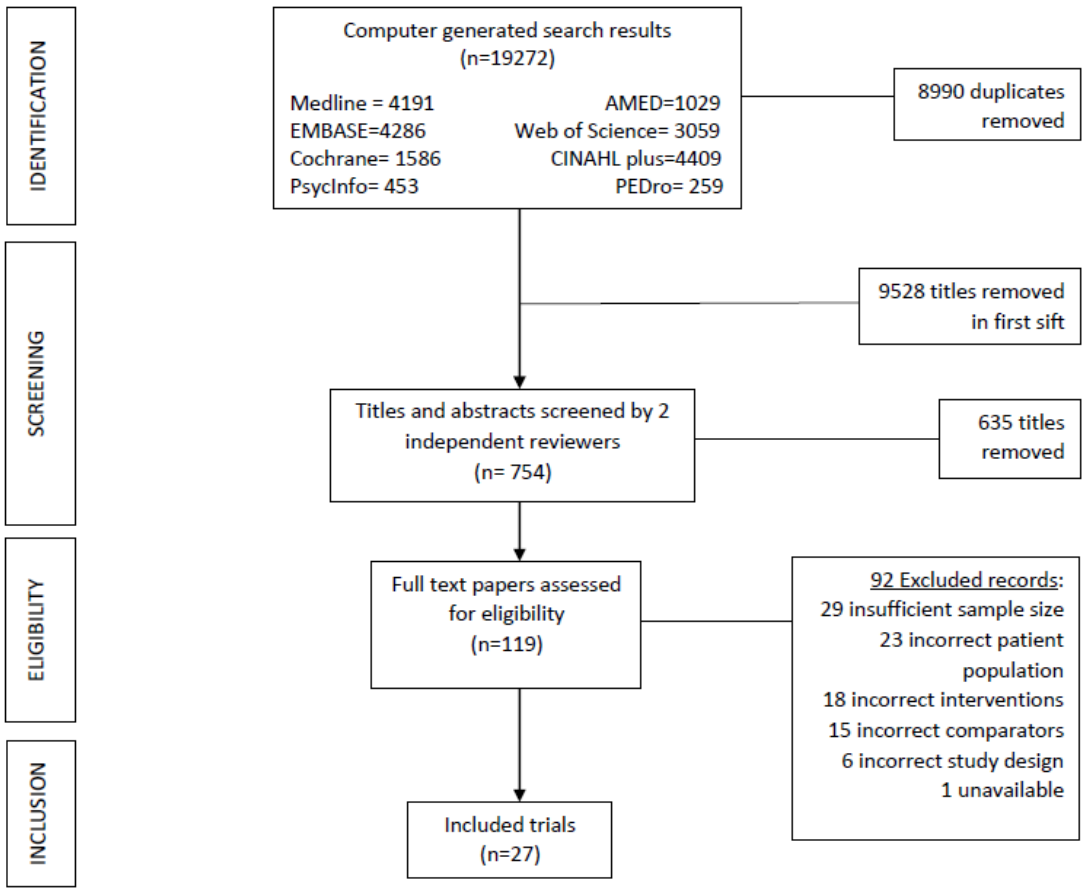
3.3.1 Search Results

The PRISMA flow diagram demonstrating the flow of information through the review is shown in Figure 3-2. After removal of duplicates, 754 remaining titles and abstracts were subsequently screened. 119 trials were identified for full-text review. 92 trials were excluded for the following reasons:

i. The trial had an insufficient sample size
ii. The trial did not include predominantly adults with persistent NSLBP
iii. The trial did not have an exercise intervention that met the eligibility criteria
iv. The trial did not have a non-exercise comparator arm
v. The study was not a RCT design
vi. The trial was not available in the UK.

A total of 27 RCTs were included.

Figure 3-2: PRISMA flow chart to represent systematic review screening and selection



3.3.2 Included RCT Characteristics

Of the 27 included RCTs, 25 were individually-randomised, and two were cluster-randomised. Sixteen were two-arm, nine were three-arm RCT designs, one had four arms, and one had six arms, tallying 69 RCT arms. The RCTs included a total of 5870 participants, giving an average of 85 participants per arm, of which 2916 were allocated to exercise, and 2954 to non-exercise control interventions. Sample sizes per RCT varied from 121 to 768. Protocols were available in the public domain for seven RCTs. Please see Table 3-2 for full detail.

Trial Setting: The 27 RCTs were from 12 different countries: six from the UK, four from the USA, three from Australia; and two each from Spain, Norway, the Netherlands and Brazil. Denmark, Hungary, Taiwan, Switzerland, Italy and Japan each led one of the included RCTs.

Table 3-2: Characteristics of Included RCTs (27 RCTs)

Trial	Trial Setting, Country	Participants (pts)	Inclusion/ Exclusion Criteria	Exercise Intervention[4]	Comparator(s)[4]	Outcomes (and Primary Time-point where specified)
Albaladejo et al. (116)	Primary care, Spain	348 pts; Mean age 51.5 yrs, duration LBP >12 wks 79.5%	Inclusion criteria: consulters for LBP, with or without referred pain. Exclusion criteria: illiteracy, bedridden, previous physiotherapy in the last 12-months, red flags, specific LBP; inflammatory rheumatologic disease; fibromyalgia or signs for suspicion of fibromyalgia.	Physiotherapy and education: Relaxation techniques, stretching and active exercises for abdominal lumbar and thoracic muscles. Standardised but intensity adapted to the patient's ability.	Comparator 1: Education and booklet on health nutrient habits and one 15-minute group talk Comparator 2: "Back book" and one 15-minute group talk	**Physical function (RMDQ) at 26 weeks (wks);** pain intensity (visual analogue scale (VAS)); catastrophising (Coping Strategies Questionnaire); health-related quality of life (Short-Form Health Survey (SF)-12)
Bronfort et al. (117)	University research clinic[4], USA	200 pts (301 total sample); Mean age 45.1±11 yrs; Median 5 yrs. duration LBP	Inclusion: 18-65yrs; mechanical LBP > 6wks duration. Exclusion: Previous lumbar fusion surgery; progressive neurological deficits; aortic or peripheral vascular disease; pain scores < 3/10; pending or current litigation; ongoing treatment.	High dose supervised low tech trunk exercise: 20 1-hour sessions over 12-weeks. High repetition core strengthening exercises with aerobic warm-up (WU).	Comparator 1: Spinal manipulation 1-2 x wk for 15-30-mins. (Comparator 2: Low dose home exercises: two 1-hr appointments followed up 1-2 wks later)[5]	**Pain (ordinal 11-box scale) at 12 wks;** Physical Function (RMDQ); health-related quality of life (SF-36); frequency of pain medication use; patient-perceived global improvement; satisfaction

[4] The trial was set at the university clinic but recruited from local community settings

[5] Where comparators or interventions in brackets were excluded from the analysis

Cambron et al. (118)	Chiropractic and allopathic clinics, USA	235 pts; Mean age 41.55 yrs.	Inclusion: >18 years, primary complaint CLBP, no contraindications to manual therapy. Tenderness over lumbar joints; forego NSAID use prior to assessment. Exclusion criteria: CNS disease, psychiatric disease or lack of cognitive ability, current and known substance abuse, not fluent or illiterate in English, morbidly obese, pregnant, currently receiving care for LBP, treated by chiropractor or PT in past 6 months, unwilling to forego other care during treatment.	Intervention: Active trunk exercise programme	Control: Flexion-distraction therapy	**Pain intensity (VAS), Physical Function (RMDQ), HRQoL (SF-36) at 4 wks;** health-care utilisation, low back biomechanics.
Cecchi et al. (119)	Outpatient physiotherapy department, Italy	210 pts; diagnosed LBP, Mean age 58.8 yrs.	Inclusion: Consulters for NSLBP; >24 weeks. Exclusion: Neurological signs; spondylolisthesis, spinal stenosis, lumbar scoliosis, rheumatoid arthritis, psychiatric disease, cognitive impairment or pain-related litigation.	Back School: 15 1-hr sessions five days a week. Education, individually tailored exercise, relaxation techniques. Group setting. (Physiotherapy: Passive and active mobilisation, active exercise and soft tissue treatment. 15 sessions 1-hr five days a week)[4]	Manipulation: Whole spine approach. 4-6 (as needed) weekly sessions of 20-minutes each for a total of 4-6-weeks (80-120-mins treatment total).	**Physical function (RMDQ) 3 wks or 4-6wks;** pain intensity (Roland Morris Pain Rating Scale);

Chen et al. (120)	Community health centre, Taiwan	127 pts; LBP; Mean age 34.2 yrs.	Inclusion: Self-reported LBP >6 months; LBP VAS >4; Exclusion: Prior surgery LBP; regular use pain medication	Stretching exercise programme 50 minutes at a time, 3 x wk, for 6-months. 10mins WU, 30-mins Back pain ex and core muscle training, 10-mins relaxation. Group setting.	Usual activities for 50-minutes at a time	Pain (VAS); Exercise Self-efficacy (Exercise self-efficacy questionnaire); 12 wks
Chown et al. (121)	Physiotherapy department, UK	159 pts (239 pts); Mean age 43 yrs.	Inclusion: LBP without leg pain > 3 months; > 18 years. Exclusion criteria: > 65 years, serious spinal disorder, main complaint of pain below hip, previous spinal surgery, additional MSK disorder, attendance or referral to a specialist pain service, medical condition, anticoagulant treatment, steroid treatment, unable to get up from floor unaided, physical therapy in previous 3 months.	Exercise group: Home stretching programme, basic postural setting, education. (Physiotherapy: Joint mobilisation, neural tension, traction, muscle imbalance, exercises)[4].Both interventions required attendance of 5 sessions over 3 months.	Osteopathy: soft tissue massage, articulations, exercise advice, education. Attendance of 5 sessions over 3-months.	**Physical function (ODI) at 6 wk and 52 wk follow up;** HRQoL (EQ-5D); fitness (Shuttle walk test).
Costa et al. (122)	Outpatient physical therapy department, Australia	154 pts, care-seeking for NSLBP; Mean age 53.7; Average duration 331.5 weeks	Inclusion: Consulters for NSLBP; >3 months duration; 18-80 yrs. age; comprehend English; motor control indicated. Exclusion: Suspected/ confirmed spinal pathology; pregnancy or lactation; nerve root compromise; Previous spinal surgery; major	Motor control: 12 30-min treatments over 8-wk period; (2 sessions a wk in first month, one session/wk in second month); 2 stages individualised motor control training	Placebo: 12 30-min treatments over 8-wk period; (2 sessions/ wk in first month, 1 session/ wk in 2 month); 20-mins detuned SWD; 5-mins detuned US	**Pain (Numeric pain rating scale (NRS)); physical function (Patient-Specific Functional Scale (PSFS)); Global impression of recovery (Global perceived effect (GPE) scale) at 8 wks**; physical function (RMDQ); recovery; recurrence LBP

			surgery scheduled; any contraindication to exercise, US or SWD			
Díaz Arribas et al. (123)	Primary care centres, Spain	137 pts; selected off PCC database with diagnosis of persistent NSLBP; mean age 39 years	Inclusion: 18-65 yrs age; NSLBP; Prescribed physio; no neurology; LBP caused by functional overload/ poor postural habits; agreed to abstain from other treatment during study; LBP 2-4 months. Exclusion: Depression; cognitive defects affecting understanding; imminent move, lack of time; contraindications to ET; red flags or yellow flags	Godelive Denys-Struyf Physiotherapy: 15 sessions of articular and muscular balancing of lumbar spine and pelvis. Mix individual and group sessions spine stabilizing ex. 2 sessions average 50-mins duration/ wk	Conventional PT: 15 treatment sessions: 14 40-min sessions of TENS plus 10-mins microwave treatment. Last session received a sheet of exercises.	**Pain (NRS) at 12 wks;** physical function (Oswestry Disability Index (ODI)); HRQoL (SF-36)
Ferreira et al. (124)	Physical therapy centres, Australia	240 pts care-seeking for CLBP; mean age 53.5yrs	Inclusion: NSLBP, 18 - 80 yrs. Exclusion: Neurological signs, specific spinal pathology or previous back surgery.	General exercise: 1-hr supervised classes with individualised exercises; 12 treatments. Motor control exercise: 12 training sessions of specific muscle stabilisation Both exercise groups delivered by a physiotherapist.	Spinal manipulative therapy: 12 sessions treatment involving joint mobilisation or manipulation techniques.	**Physical function (PSFS), global perceived effect (11-point scale) at 8 wks;** pain (VAS) and disability (RMDQ).
Garcia et al. (125)	Physiotherapy clinic, Brazil	148 pts, care-seeking for CLBP; Mean age 76; Average duration 42 months	Inclusion: >3/10 Pain, 18-80yrs, read Portuguese. Exclusion: Contra-indication to exercise/ electrotherapy; nerve root compromise, serious	Exercise: 10 sessions, twice weekly 5-wks; 30-40-mins long. Repeated ex in prescribed direction + education.	Placebo: detuned US 5-min; SWD detuned 25-min, +educational booklet "the back book"	**Pain intensity (NRS); physical function (RMDQ) at 5 wks;** (PSFS); kinesiophobia (Tampa Scale of Kinesiophobia (TSK));

			spinal pathology, CVD/ metabolic disease, previous back surgery or pregnancy	Classified into syndrome. 10-15 reps 3-5x day and educational booklet "the back book" and "treat your own back"		global impression recovery (GPE)
Goldby et al. (126)	Physiotherapy Department, UK	213 pts, referred for physiotherapy; mean age 42yrs; average duration 11.72 ±9.85yrs	Inclusion: NSLBP >12 weeks, 18-65yrs, literate English. Exclusion: Non-mechanical LBP; stenosis/ spondylolisthesis, worsening/ significant neurological deficit; inflammatory joint disease; lower limb pathology; metastatic disease; medically unable to participate in exercise; Chronic Pain syndrome; anxiety/ neurosis; pregnancy	Exercise: 10-wk spinal stabilization. Staffed by two physiotherapists <12 patients /1-hour class with back school	(Physiotherapy: SMT; no spinal stabilisation exercises included electrotherapy and any other exercise. Maximum ten sessions + Back school)[4] Minimal intervention. Booklet "Back in action". Back school.	Pain (NRS); physical function (ODI), Handicap (Low back outcome score) and Impairment (lumbar flexion and timed walking test), Quality of life and dysfunction (Nottingham Health Profile) at 12 wks
Groessl et al. (127)	Veterans Affairs Medical Centre, USA	150 pts referred by clinicians, mean age 53.4 yrs.	Inclusion: Aged >18 yrs, NSLBP >6 months, English literacy, no new pain treatments in last month, willing to attend yoga or be assigned to delayed treatment, willing to not change treatments unless medically necessary. Exclusion criteria: recent back surgery (past 12 months), systemic causes of back pain, morbid obesity, acute nerve root	12-wk Yoga intervention: 2 60-min instructor-led sessions a wk. Encouraged to do home exercises 15-20-minutes on non-instructor led days.	Delayed yoga after 12-months.	**Physical Function (RMDQ) at 12 wks;** pain intensity (Brief pain Inventory); non-study treatments and medications – use.

			compression, chronic lumbar radicular pain, serious unstable comorbidities or psychiatric conditions, positive Romberg's test, practiced yoga in past year.			
Hansen et al. (128)	Workers for Airline Company[6], Denmark	121 pts (180 total sample) self-reported LBP; Mean age 41.1yrs; average duration 27 weeks	Inclusion: Persistent LBP (50%)>3 months; sub-chronic > 4 weeks and 2 pain episodes /month in past 12 months. Exclusion: Permanently disabled; specific causes of LBP including root compression; pregnancy or lactation excluded; Previous spinal fusion	Intensive dynamic back muscle training: 3 exercises performed five series 10x ea.; 1-min break. After 10-min rest repeated all ex with 300 contractions over 4-weeks	(Standard PT: soft tissue treatment, flexibility ex for lumbar spine and pelvis, coordination ex, counselling, isometric back ex)[4] Placebo: 20-mins semi-hot packs, traction intermittent 20 mins	Pain scores (interval scale) at 4 wks; overall treatment effect
Hall et al. (129)	General community, Australia	160 volunteers, recruited from advertisements with moderate pain/ disability; Mean age 44.4±13.2yrs	Inclusion: minimal /moderate pain or moderate activity limitation reported; NSLBP+ leg pain Exclusion: suspected/ known spinal pathology; contraindications to exercise; scheduled for spinal surgery.	Tai chi sessions 40 mins long; 18 sessions over 10-wks, group setting (2 x wk for 8-wks then 1 x wk for 2-wks)	Waitlist controls	**Bothersomeness of pain (NRS)at 10 wks;** physical function (ODI; Pain disability index (PDI); Quebec Back Pain and Disability Scale (QBPDS); PSFS), GPE
Harris et al. (130)	Clinics, Norway	159 pts (214 total) sick-listed patients contacted by Norwegian	Inclusion: sick leave for LBP 2-10 months Exclusion: not fluent Norwegian; age >60; pregnancy; awaiting	Brief intervention + physical exercise: 3 x wk at clinic for 3-months. 90-min sessions group.	(Brief intervention+ CBT group: 7 group sessions over 3-months including graded activity.	**Increased work participation at 52 wks (change in sick leave status);** physical function (ODI), anxiety &

[6] Staff from a large organisation volunteered following an advertisement in an internal company newspaper.

		Labour and Welfare Administration; Mean age 44.8yrs.	surgery; disabled; comorbidities; psychological conditions; insurance trials.	Individual goals set and adaptation of programme accordingly. Strength, endurance and relaxation in group setting.	Psychiatrist led and lasted 90min sessions.)[4] Brief intervention: 2 sessions over five days with option of 2 booster sessions	depression (Hospital Anxiety and Depression Score (HADS)), subjective health complaints (Subjective health complaints index), coping (Utrecht Coping List), fear-avoidance (Fear-avoidance beliefs questionnaire (FABQ))
Hildebrandt et al. (131)	GP practices, The Netherlands	222 pts Previous LBP care-seekers	Inclusion: 18-55yrs age; NSLBP, >4 episodes in past one year. Exclusion: prev. Cesar or Mensendieck therapy; other psychological problem for which receiving medication; understand/ communicate in Dutch.	Maximum 18 sessions of 35-minutes, two sessions/ wk for 3-months. No other interventions (physiotherapy/ medication)	Guideline managed LBP No other interventions (physiotherapy/ medication)	Self-reported recovery Improvement of posture (thoracic, lumbar spine, pelvis) at 12 wks
Jans et al. (132)	Cesar Therapists, The Netherlands	201 pts; mean age 35 years	Inclusion: 18-55 yrs.; NSLBP; pain> 6 weeks or minimum of 4 episodes in previous 12 months; not had previous intervention; No comorbidities; able to complete questionnaires in Dutch; no other care for LBP.	Both groups received treatment for 3-months 10-14 sessions, follow up; intervention group had 6 follow up treatments in 6-month period	Both groups received treatment for 3-months (10-14) sessions. No further intervention	Self-reported recovery from LBP at 26 wks; physical function (QBPDS); intensity and duration of LBP (Van Korff Pain scale), general health (interval scale)
Járomi et al. (133)	Local academic medical centres, Hungary	137 care-seeking pts; mean age 41.4 years	Inclusion: Diagnosed with persistent NSLBP, employed in health-care setting, worked as bedside nurses for >3 years.	Back school: twice weekly for 12-weeks in 60-min session each. Educational component and exercise programme (consisting of muscle	Control: Brief written lifestyle guidance	**Lifting technique analysis (Zebris WinSpine Triple Lumbar biomechanical motion analysis); Pain intensity over the past week (VAS) at 12 wks.**

			Exclusion: One or more diagnoses: acute/ subacute LBP, Specific or sinister causes of back pain; surgical recommendations, tumour, failed back syndrome, rheumatic or musculoskeletal disease, depression and other psychiatric disease, chronic pain syndrome, gynaecological or urological conditions, pregnancy, receiving current or recent physiotherapy, previous spinal surgery in past 6-months.	strengthening, correct posture, mobilising and stretching activities, and practice of proper patient handling).		
Johnson et al. (134)	Family medical practices, UK	196 pts, consulting with LBP; mean age 47.9yrs	Inclusion: Consulting for persistent NSLBP Exclusion: previous consulting for LBP in past 6-months; "red flags", pregnancy or recent childbirth, comorbidities that may prevent full participation; previous spinal surgery, major psychiatric disorder diagnosed or under investigation, history of drug or alcohol abuse in past 5 years.	Both groups mailed educational pack including a book and audio-cassette regarding self-management advice. Eight 2-hour group sessions over 6-weeks including exercise and changing beliefs and thoughts.	Both groups mailed educational pack including a book and audio-cassette regarding self-management advice. Treatment as usual by GP.	**Pain severity (VAS); physical function (RMDQ) at 52 wks;** HRQoL (Euro-Qol 5D (EQ-5D)).

Maul et al. (135)	University hospital staff[7], Switzerland	148 hospital employees reporting LBP; Mean age 39 yrs; average duration pain 12 yrs	Inclusion: >30d LBP in past 1 yr.; 8-30 days LBP +reported disability in daily tasks >12 months; 20-55 yrs. age; read and write in German/ Italian Exclusion: CVD/ metabolic disease; progressive radicular neurological deficit; inflammatory disease of spine; prior spinal surgery; regular strength training in prior 6 months; pregnant women	Low back school as per control. 3-month programme with three phases of individual training with a duration of 4 -wks each (attendance 1-hr x 2/wk).	Low back school: 3 sessions 1-hour each	Lifting capacity (PILE test) at 12 wks; isokinetic trunk strength; ROM Lumbar spine flexion, extension and side bending; Isometric muscle endurance; aerobic capacity (box step); frequency and duration LBP (Nordic questionnaire); physical function (RMDQ and Waddell Q); Pain intensity (NRS, analgesic intake, Quantitative pain drawing); Pain characteristics (McGill q); General well-being (General well-being questionnaire); General beliefs (sense of coherence questionnaire)
Miyamoto et al. (136)	Primary care, Brazil	295 pts voluntary recruitment; mean age 47.9 yrs; average duration symptoms 45.75 months	Inclusion criteria: NSLBP >3months. Exclusion criteria: Pilates treatment for LBP in prev 3-months, serious spinal pathologies, nerve root compromise, previous or scheduled spinal surgery, contraindications to exercise, pregnancy	Three different groups of Pilates: • once a wk • twice a wk • three times a wk All individualised, one to one sessions for 6-wks.	Booklet group, advice only: receive pilates at 12-months.	**Pain Intensity (NRS); Physical function (RMDQ) at 6 wks;** global perceived effect (GPE), physical function (PSFS); catastrophising (pain catastrophising scale); kinesiophobia (TSK); HRQoL (SF-6D)

[7] Staff volunteered and were invited further to questionnaires

Moffett et al. (137)	Primary care, UK	219 pts (315 total sample). Mean age 44yrs, 67% chronic pain.	Inclusion criteria: >18yrs with back pain or neck pain >2 weeks' duration, considered to be of non-systemic origin. Exclusion criteria: patients with a score < 4 RMDQ or < 10 on the Northwick Park Neck Pain Questionnaire. Previous physical therapy within the previous 3 months or planning to use private physical therapy alongside the NHS physiotherapy. Referral from a hospital consultant, possibility of serious pathology or pregnancy.	McKenzie treatment: exercise according to directional preference.	SFA: Solution finding approach: based on cognitive behavioural principles. Reassurance, goalsetting, assessment, explanation about the condition.	**Physical Activity Avoidance subscale of the TSK**, **and physical function (RMDQ or Neck Pain questionnaire) at 6 wks**. Health control (Multidimensional Health Locus of Control), HRQoL (SF-12), (EQ-5D); self-efficacy (Pain Self-Efficacy Questionnaire), anxiety and depression (HADS).
Russell et al. (138)	NHS and Private centres across UK	768 pts (1334 total sample); mean age 43.1 yrs; 58.7% had back pain episode lasting more than 90 days	Inclusion: 18-65 yrs; consulted with simple LBP; score of >4 on RMDQ. Agree to avoid physical treatments for 3 months. Exclusion: Possibility serious spinal disorder; pain mostly below the knee; previous spinal surgery; another MSK problem more troublesome; attended or referred to a pain management clinic; severe psychiatric or psychological disorder; another medical condition	Exercise Programme: Back to Fitness. Initial individual assessment followed by group classes incorporating CBT principles. Based in local communities. Up to 8 60-minute sessions over 4-8-weeks and a "refresher" class 12-weeks after randomisation.	Best Care in general practice: "active management" of back pain, copies of Back Book. Spinal Manipulation: Up to 8 20-minute sessions either in NHS or in private rooms over 12-weeks. (Combined group: Combination 6-weeks SMT followed by 6-weeks of exercise classes	**Physical function (RMDQ) at 12 wks; pain** (modified van Korff scales); general beliefs (Back Beliefs Questionnaire); fear-avoidance beliefs (FABQ); HRQoL (SF-36; EQ-5D).

			that could interfere with therapy; moderate to severe hypertension; taking anticoagulants/ long term steroids. Unable to walk 100m, get up and down from floor unaided. Previous physiotherapy in past 3 months; RMDQ score <3 ; unable to read or write fluently in English		and a refresher class)[4]	
Saper et al. (139)	Community health centres, USA	191 pts (320 total sample) Mostly Care seeking individuals; Mean age 45.7 years	Inclusion: persistent NSLBP; 18-64; pain > 4/10 in previous week; English speaking Exclusion: specific causes LBP	Yoga 12 weekly 75-min classes; 30-min HEP; maintenance classes	(PT treatment based classification, graded ex. 15 60-min appts over 12-wks; FABQ >29 provided Back Book + education)[4] Brief Intervention: The back pain help book; 3-wks newsletter and brief check-in call 6 wks	**Physical function (RMDQ); Pain (NRS) at 12 wks;** Pain medication use (self-report) global improvement; patient satisfaction; HRQoL (SF-36)
Shirado et al. (140)	Outpatient clinics, Japan	201 pts; Consulted orthopaedic surgeons for NSLBP; mean age 42.2y	Inclusion: 20-64yrs; persistent NSLBP; no neurological deficit; SLR >70deg, negative femoral nerve stretch test, muscle strength> 4/5. Exclusion: red flag causes LBP, previous spinal surgery. No comorbidities.	Trunk muscle strengthening and stretching ex. Demonstrated by HCP. 10 x 2 sets at Home. 1st class 15-30-mins to teach, then visit 1-2x wk x 8 wks supervised exercise + NSAIDs. (group setting)	NSAIDs (loxoprofen sodium/ diclofenac sodium/ zaltoprofen) x 3 + rebamipide/ sodium azulenesulfonate/ teprenone	**Pain (VAS); physical function (RMDQ); HRQoL (Japan LBP evaluation questionnaire) at 8wks;** Finger floor distance

Storrø, Moen and Svebak, (141)	Local fitness centres, Norway	218 pts identified by local national insurance office from GP diagnosis; mean age 44yrs.	Inclusion: GP diagnosis of non-specific neck and shoulder pain, LBP or LBP with radiation. Time since being sick-listed >4<12 months.	Groups met 3 x wk for 2-hours over 4-wks with medical doctor, psychologist and physiotherapist. Tailored physical exercises provided.	Treatment as usual – referral to physiotherapy/ chiropractor etc.	**Sick listed or not at 52 wks.**
Tilbrook, Cox and Hewitt, (142)	Nonmedical centres[8], UK	313 mixed consulters and non-consulters LBP; mean age 46.4 yrs.; average duration pain 84 months	Inclusion: >4 RMDQ; consulted for LBP in past 18 months; ability to attend 1 venue. Exclusion: yoga in prior 6 months; unable to get off floor unaided or use stairs; pregnancy; life-threatening comorbidities; previous spinal surgery; severe psychiatric problems /alcohol dependency; indications serious spinal neurological abnormality (such as cauda equina or cord compression)	Yoga: 12 classes gradually progressed class over 3-months. 75-mins long; encouraged to practice at home at least 2 x week.	Back pain education booklet with usual care; one-time session of yoga after final follow-up	**Physical function (RMDQ) at 12 wks;** HRQoL (SF-12); Back pain scores (Aberdeen Back Pain scale); pain self-efficacy (PSEQ); HRQoL (EQ-5D); number days in bed and days of restricted activity; medication use and health-care usage; beliefs, expectations and preferences for treatment; use of yoga at home

Bold outcome domains signify specified primary outcome domains, measures and primary follow-up points. Abbreviations used: yrs – years; wk(s)- week(s); min- minute; MSK – musculoskeletal; SMT- spinal manipulation therapy; CBT – Cognitive behavioural therapy; HCP – Health-care professional; GP- General practitioner; pts –participants; US –Ultrasound; SWD – shortwave diathermy; CVD- cardiovascular disease; ex- exercise; HEP- home exercise programme; PT- physiotherapy; appts- appointments; NSAIDs – nonsteroidal anti-inflammatory drugs; WU – warmup; C1/2: control group ½; ROM – range of movement; q - questionnaire

[8] Yoga was delivered in nonmedical settings although patients were recruited from GP surgeries

Seven RCTs included an exercise arm with a mixed exercise component which included a regime of stretching and strengthening exercise and/or education. Four RCTs had a strengthening focussed exercise arm, five focussed on spinal stabilisation and motor control exercise, and two trials had a stretching focussed exercise arm. Three RCTs tested yoga (127,139,142) as the exercise intervention, one Pilates (136), and one tested Tai Chi (129). Five RCTs tested a specific exercise therapy approach such as McKenzie (125,137), Godelieve Denys-Struyf (123) or Cesar exercise therapy (131,132). Most RCTs utilised a group exercise format (n=18), fewer trials (n=9) delivered their exercise intervention in a one-to-one format. Only one RCT used a mix of both group and one-to-one formats (123). Home exercise programmes (HEP) were prescribed in 12 RCTs (see Table 3-3). Ferreira et al. (124) included two exercise arms, and Miyamoto et al. (136) included three exercise arms so in total across the 27 RCTs, there were 30 exercise arms. The dosage (including frequency and duration) of exercise tested in each RCT varied greatly, as seen in Table 3-3, with the average exercise intervention comprising one hour, twice per week over ten weeks (thus, a high dosage programme as described by Hayden et al. (143)). However, only six RCTs delivered a high dosage intervention (117,120,123,127,130,133). The intensity of exercise interventions was generally not described; thus, these data were not possible to extract.

Comparators[9]*:* Fourteen RCTs (116,120,139,141,126,130–135,138) used a minimal intervention (no exercise) control group, six RCTs (117–

[9] These figures do not add to the 27 as two trials had more than one comparator arm

119,121,124,138) compared exercise to spinal manipulation, four RCTs (127,129,136,142) had waiting list control groups, three RCTs (122,125,128) compared their exercise intervention to a placebo intervention (see Table 3-2 for descriptions of the placebo interventions used), and two RCTs (123,137) compared exercise to other forms of usual care physiotherapy (excluding exercise) (n=29 comparator arms).

Follow-up End-point: The primary end-point varied across RCTs from 3 weeks (119) to 12 months (130,134,141), with an average follow-up time of 15.1 weeks.

Table 3-3: Variation in exercise dosage across included RCTs

Trial	Delivered by	Length of session	Frequency	Duration	HEP
Albaladejo et al. (116)	Physiotherapist	1 hour	4 days	N/a	Yes
Bronfort et al. (117)	Exercise therapist	1 hour	2 x week	12 weeks	No*
Cambron et al. (118)	Physiotherapists	Unclear	2-4 x week	4 weeks	No
Cecchi et al. (119)	Physiotherapist (Individual)	1 hour	5 x week	3 weeks	No
Chen et al. (120)	Expert in sports, health and leisure	50 mins	3 x week	6 months	No
Chown et al. (121)	Physiotherapist (Group)	30 mins	5 sessions	12 weeks	Yes
Costa et al. (122)	Physiotherapist	30 mins	1-2 x week	8 weeks	Yes
Díaz Arribas et al. (123)	Physiotherapist	50 mins	2 x week	8 weeks	No
Ferreira et al. (124)	Physiotherapist (Group)	1 hour	12 sessions	8 weeks	Yes
	Physiotherapist (Individual)	1 hour	12 sessions	8 weeks	Yes
Garcia et al. (125)	Physiotherapist	30-40mins	10 sessions	5 weeks	Yes
Goldby et al. (126)	Physiotherapist	1 hour	10 sessions	10 weeks	No
Groessl et al. (127)	Yoga instructor	1 hour	2 x week	12 weeks	Yes
Hansen et al. (128)	Unclear (likely Physiotherapist)	1 hour	2x week	4 weeks	No
Hall et al. (129)	Tai Chi Instructor	40 mins	18 sessions	10 weeks	No
Harris et al. (130)	Physiotherapist	90 mins	3 x week	12 weeks	No
Hildebrandt et al. (131)	Cesar exercise therapist	35 mins	2 x week	12 weeks	No
Jans et al. (132)	Cesar exercise therapist	35 mins	1 x week	12 weeks	No
Jaromi et al. (133)	Unclear	60 mins	2 x week	12 weeks	Yes
Johnson et al. (134)	Physiotherapist	2 hours	8 sessions	6 weeks	Yes
Maul et al. (135)	Physiotherapist	1 hour	1-2 x week	12 weeks	No
Miyamoto et al. (136)	Physiotherapist	1 hour	1-3 x week	6 weeks	No
Moffett et al. (137)	Physiotherapist	Unclear	4 sessions	Unclear	No
Russell et al. (138)	Physiotherapist	1 hour	2 x week	4 weeks	No
Saper et al. (139)	Yoga Instructor	75 mins	1 x week	12 weeks	Yes
Shirado et al. (140)	Nurse/ physiotherapist	15-30 mins	1-2 x week	8 weeks	Yes
Storrø, Moen and Svebak, (141)	Multidisciplinary team	2 hours	3x week	12 weeks	No
Tilbrook, Cox and Hewitt, (142)	Yoga instructor	75 mins	1 x week	12 weeks	Yes

HEP stands for home exercise programme. *Bronfort et al. (2011) did not include a HEP in their included exercise arm but did include one arm in their RCT that used HEP only.

3.3.3 Characteristics of Excluded Trials

The flowchart (Figure 3-2) summarises the characteristics of the 92 excluded studies. Further detail is provided in Appendix 9b: Excluded Studies from the Systematic Review with Reasons.

3.3.4 Risk of Bias Assessment

There was substantial variation in the overall risk of bias across the 27 RCTs. The summary of the risk of bias assessment in Figure 3-3 and 3-4 shows that only 22% of RCTs (n=6) demonstrated overall low risk of bias with five of the risk of bias domains met. Most RCTs described adequate sequence generation (n=22/27). No RCTs had low risk of bias for blinding, despite trial-specific measures specifically described by Costa et al. (122). The high risk due to inadequate blinding was agreed by the reviewers due to the nature of participating in an exercise intervention in comparison to a non-exercise control. 37% (n=10) of RCTs had a low risk of attrition bias. 78% (n=21) of RCTs had low risk of bias for selective reporting, whilst 52% (n=14) were at risk of other biases.

Figure 3-3: Summary of the risk of bias of the 27 included RCTs

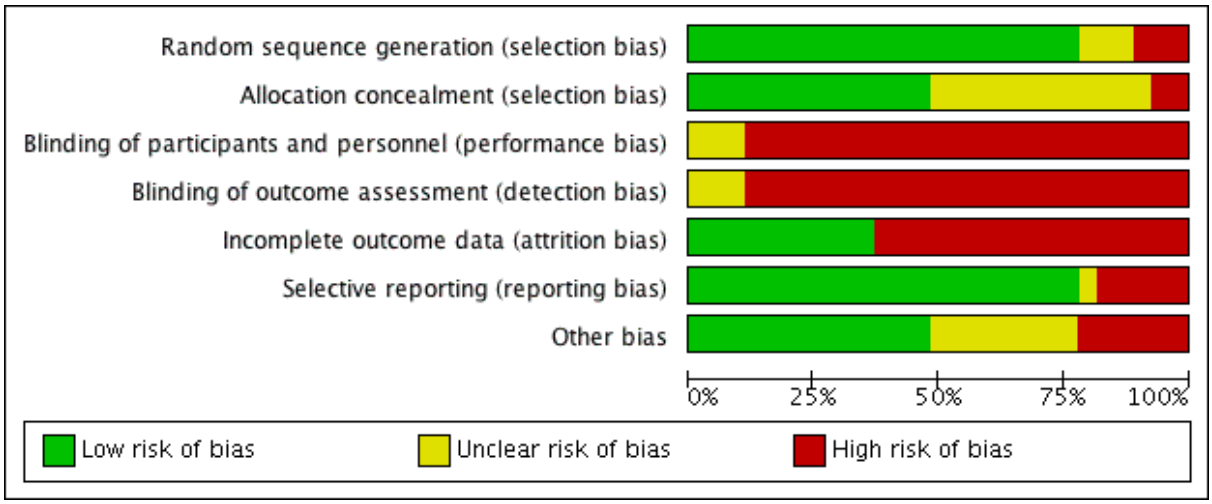
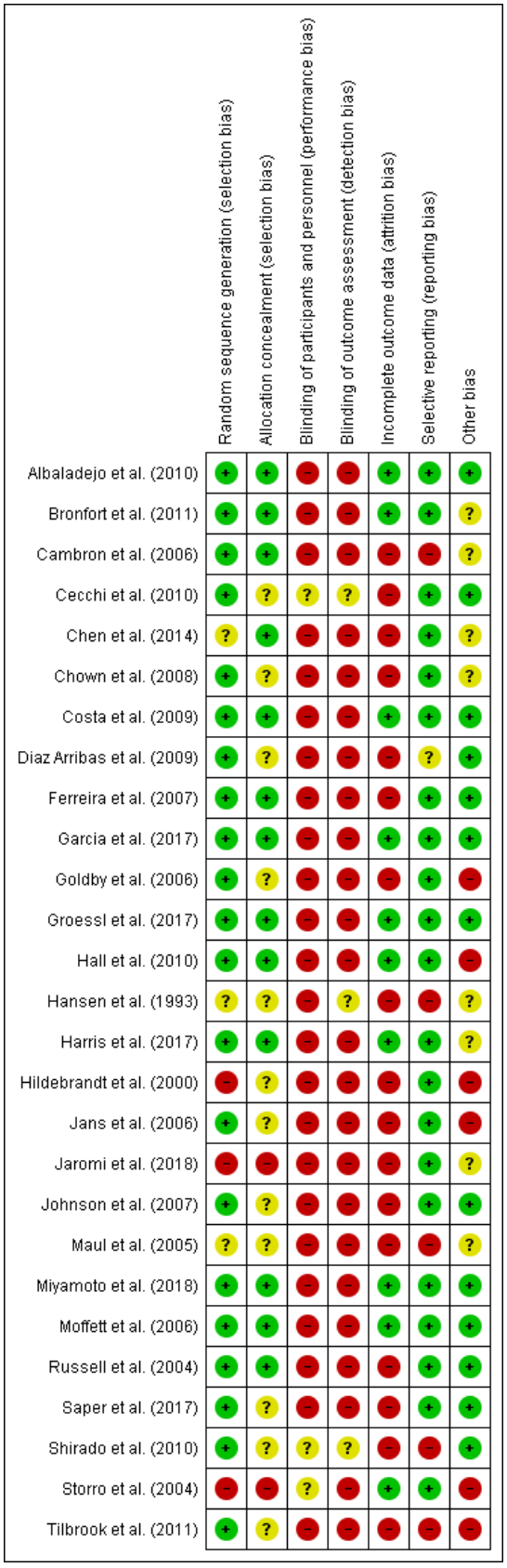


Figure 3-4: Summary of the risk of bias components of included RCTs



3.3.5 Descriptive Analysis

Treatment targets, outcome domains, and outcome measures were extracted from all included 27 RCTs. Further information is provided in Appendix 9c: Extracted Treatment Targets and Outcomes of Included Trials; Appendix 9d: Extracted Treatment Targets of Included Trials; Appendix 9e: Extracted Secondary Outcome Domains and Measures of Included Trials.

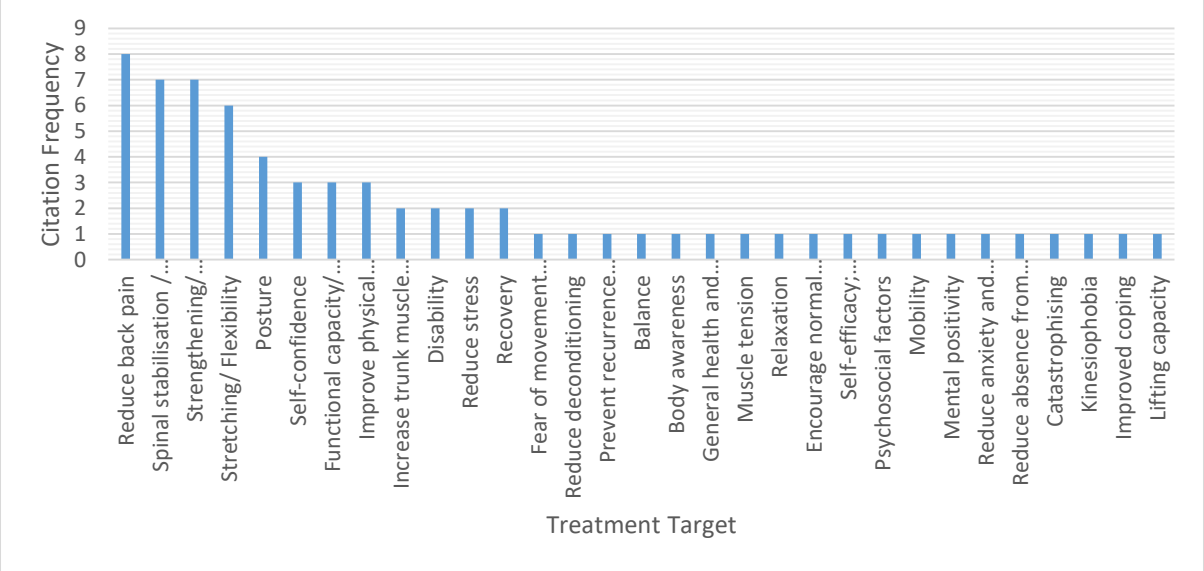
3.3.5.1 **Treatment Targets**

In total, 18 of the 27 RCTs explicitly specified treatment targets of the exercise intervention that they investigated, such as “the goal was to address fear-avoidance and movement-phobia and help to re-establish normal movement patterns” (130). Although Cecchi et al. (119) specified a treatment target relating directly to one of their exercise arms, they did not specify a treatment target for their other exercise intervention arm. Six RCTs did not specify treatment targets but alluded to these in the background section of their papers, resulting in “inferred” treatment targets (see Appendix 9 c: Extracted Treatment Targets and Outcomes of Included Trials). For example, Hansen et al. (128) did not state any specific targets of their exercise intervention but mentioned "tense back muscles being stretched through exercise", "bodybuilding", and "hyperextending back exercises" in the background section of their paper. Three RCTs did not specify or allude to any exercise treatment targets (116,121,132).

In total, 31 different treatment targets were identified in these 18 RCTs, with a range of 0 to 7 targets specified or alluded to per RCT (median of 3), but 19 of

these were each only mentioned in one RCT. These were extracted verbatim and grouped together if they related to similar constructs (such as 'reducing kinesiophobia' and 'reducing fear of movement'). However, where the treatment targets were felt to be different, constructs were left as separate items (such as coordination, posture and normal movement). The most frequently reported treatment targets were 'reducing back pain' (9 trials), 'increasing muscle strength' (8 trials), 'targeting spinal stabilisation or altered spinal control or trunk stability' (8 trials), 'stretching or improved flexibility' (7 trials) and 'improved posture' (4 trials). Figure 3-5 shows the frequencies of the stated exercise treatment targets highlighting how many were mentioned in only one RCT. More detail is provided in Appendix 9d: Extracted Treatment Targets of Included Trials.

Figure 3-5: Frequencies of stated exercise treatment targets in included RCTs



3.3.5.2 **Outcome Domain and Outcome Measure Data Extraction**

Eleven RCTs specified one primary outcome domain, seven RCTs specified two primary outcome domains, and three RCTs specified three primary

outcome domains. Six RCTs did not explicitly identify any primary outcome domain, and therefore, the first-mentioned outcome domain was taken to be their primary outcome domain (120,126,128,131,135). Six primary outcome domains were identified, of which the most frequent were physical function (n=15), pain (n=14), and global perceived effect (including patient-perceived recovery) (n=6). The outcome measures used in the six domains are summarised in Table 3-4.

Table 3-4: Primary outcome domains and measures (where all designated primary outcome domains have been extracted and first mentioned outcomes in those not designated)

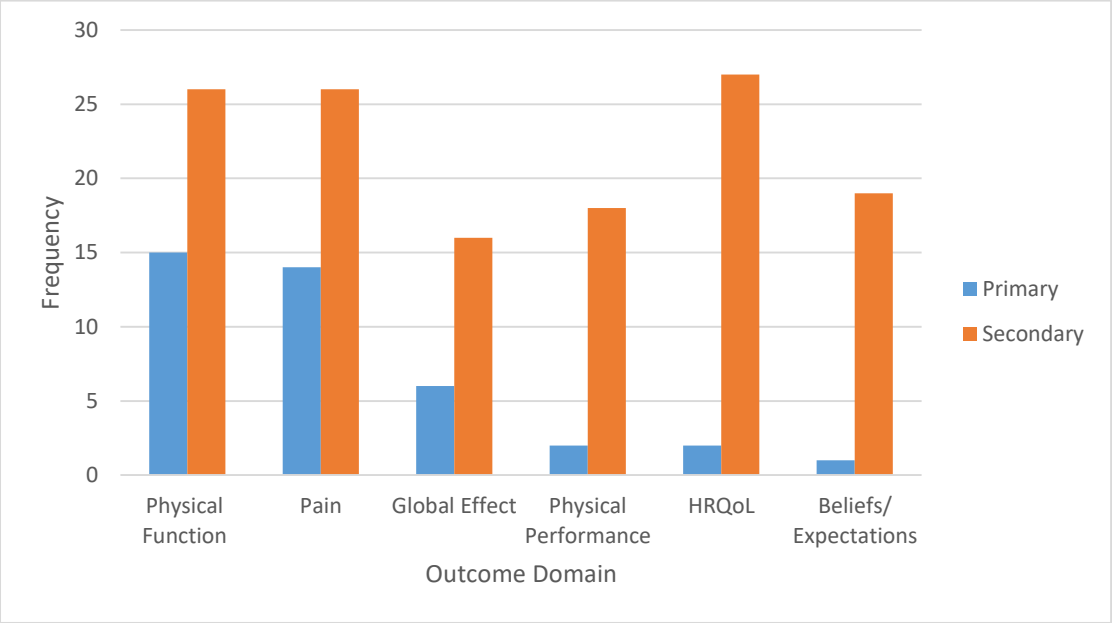
Primary Outcome Domain	**Number of RCTs (n)**	**Outcome Measure**	**Number of RCTs (n)**
Physical Function	15	Roland and Morris Disability Questionnaire	12
		Patient-Specific Functional Scale	2
		Oswestry Disability Index	1
Pain	14	Numeric Rating Scale	8
		Visual Analogue Scale	5
		Ordinal 11-box scale	1
Global Perceived Effect	6	Global Perceived Effect	2
		Self-reported recovery (7-point scale)	2
		Increased work participation	2
Physical Performance	2	Lifting capacity (PILE test); Analysis of motion (quantitative and qualitative)	2
Health-Related Quality of Life	2	Japan Low Back Pain Evaluation Questionnaire	1
		Short Form-36	1
Beliefs and Expectations	1	Tampa Scale of Kinesiophobia	1

The number of RCTs does not total 27 as some specified more than one primary outcome domain

The most frequently cited secondary outcome domain was HRQoL (n=27) followed by pain and physical function (n=26 each). In total, 135 secondary outcome measures were also extracted across the 27 included RCTs and grouped per domain. On average, 5 secondary outcome measures were used

per RCT, with a range from 0 to 14. Figure 3-6 summarises the frequency of primary and secondary outcome domains across the included RCTs, with more detail in Appendix 9e: Extracted Secondary Outcome Domains and Measures of Included Trials.

Figure 3-6: Frequency of outcome domains in included RCTs



HRQoL is an abbreviation for health-related quality of life

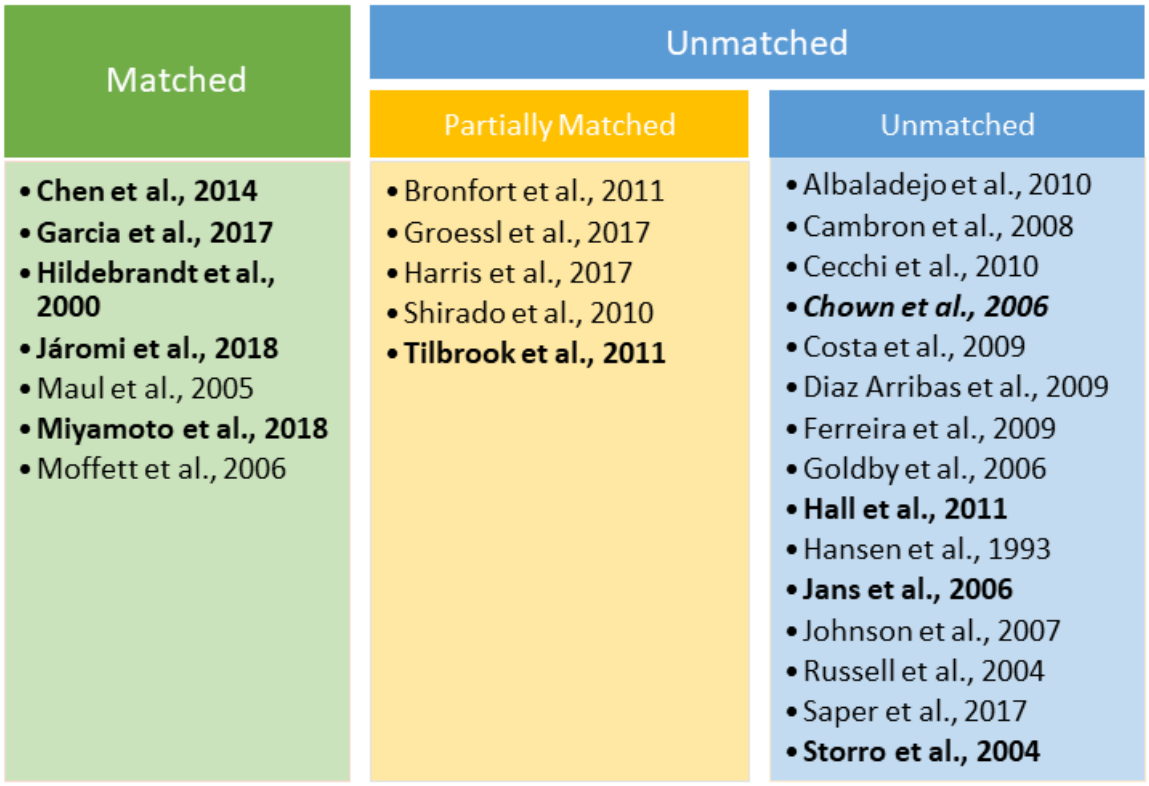
3.3.6 Categorisation of Matching Status

RCTs were allocated into groups representing ‘matched’ or ‘unmatched’ primary outcome domains and treatment targets through discussion and agreement by the reviewing team. The ‘partial matched’ sub-group (within the ‘unmatched’ group) was useful to identify RCTs for the secondary data analyses later in this thesis (chapters 4 and 5). Still, as the primary outcome in the ‘partial matched’ group was also unmatched, they were grouped with the ‘unmatched’ trials for this

stage of research. For more detail regarding the extracted data, please see appendices 9: c-e.

Figure 3-7 shows that seven RCTs (25.9%) were classified as 'matched'. Most of the RCTs (n=20, 74.1%) included in this review could not be classified as having matched outcome domains with their exercise treatment targets. The number of RCTs with statistically significant and clinically important findings is displayed in Figure 3-7 (bold), where the proportion of statistically significant findings (p<0.05) is highest in the matched category (71%).

Figure 3-7: Categorisation of RCTs according to matching status



Items in bold indicate statistically significant (p<0.05) results in favour of the superiority of the exercise group for the primary outcome domain, bold and italics indicate statistically significant (p<0.05) results in favour of the control arm.

3.3.6.1 **Characteristics of Categorised Groups**

3.3.6.1.1 ***Matched Category***

In total, 1197 participants were included in seven RCTs judged to be matched (i.e. the primary outcome matched at least one of the trial team's stated exercise treatment targets). Table 3-5 provides a summary of the details of these seven matched RCTs. They reported a wide variety of exercise treatment targets and outcomes. Each RCT had a different primary outcome, although three had nominated two primary outcomes (both Garcia et al. (125) and Miyamoto et al. (136) specified pain and physical function, whereas, Moffett et al. (137) specified fear of movement and physical function). The other four RCTs did not specify primary outcomes. An average of 5.1 (range 2-12) outcome measures were reported in each of the seven RCT papers.

Garcia et al. (125) and Moffett et al. (137) described their exercise treatment targets in the methods, and Miyamoto et al. (136) described their targets in the background section of their papers. The other four RCTs implied treatment targets in the introduction sections of their papers. For example, Chen et al. (120) referred in detail to pain and self-efficacy as treatment targets in their introduction, and they measured these two outcome domains only.

Statistically significant findings in favour of exercise therapy in the nominated or designated primary outcome domain were noted in five of the seven RCTs (bold in Figure 3-7). However, clinically meaningful results (greater than the MCID) were only noted in two RCTs: in Miyamoto et al. (112) between one of the

exercise intervention arms (Pilates twice weekly) versus usual care, in both primary outcomes, as well as in Jaromi et al. (133) in both primary outcomes in favour of the exercise arm.

Table 3-5: Summary of matched RCTs

Trial	**Recruitment Strategy (Consulters/ Non-consulters)**	**Exercise Intervention Type**	**Comparator Group**	**Number of Outcomes Used**	**Specified / Inferred Treatment Targets**
Chen et al. (120)	Non-consulters	Stretching and strengthening	Usual activities	2	Inferred
Garcia et al. (125)	Consulters	McKenzie	Back School	3	Specified
Hildebrandt et al. (131)	Consulters	Cesar Therapy	GP care	2	Inferred
Járomi et al. (133)	Consulters	Back school	Brief intervention	2	Inferred
Maul et al. (135)	Non-consulters	Strengthening	Back School	12	Inferred
Miyamoto et al. (136)	Non-consulters	Pilates	Brief Intervention	8	Specified
Moffett et al. (137)	Consulters	McKenzie	Solution Finding Approach Physiotherapy	7	Specified

3.3.6.1.2 *Unmatched Category*

Twenty RCTs with 4510 participants were judged to be ‘unmatched’ (i.e. the primary outcome was not matched to any identified treatment targets), and the extracted characteristics are summarised in Table 3-6. These RCTs predominantly assessed pain as the primary outcome (58%), three assessed pain and physical function as primary outcomes (118,134,139). One trial (122) assessed pain, physical function and recovery, and another RCT (140) specified pain, physical function and HRQoL all as primary outcomes. Four RCTs (20%) used physical function as the primary outcome (119,121,127,142), and one further used return to work (5%) (130). An average of 4.8 outcome measures was used per RCT (range 1-10).

The exercise treatment targets of fourteen RCTs were specified in the published results papers, apart from Groessl et al. (127) who specified theirs in the published protocol paper (144). Four RCTs implied their exercise treatment targets in the introduction sections of their papers, and three RCTs did not mention any targets at all. Only four (20%) unmatched RCTs showed statistically significant results in favour of exercise therapy in the primary outcome (bold in Figure 3-7) – and one RCT showed statistically significant results in favour of the control group (bold and italics in Figure 3-7). Both the trials by Tilbrook et al. (142) and Hall et al. (129) demonstrated a clinically meaningful difference (greater than the MCID) in favour of exercise versus the control arm using their primary outcome.

Table 3-6: Summary of unmatched RCTs

Trial	Recruitment Strategy (Consulters/ Non-consulters)	Exercise Intervention	Comparator Group	Number of Outcomes Used	Specified / Inferred Treatment Targets
Alabaladejo et al. (116)	Consulters	Stretching and Strengthening	Education and Booklet	4	None
Bronfort et al. (117)	Non-consulters	Strengthening	Manual Therapy	6	Specified
Cambron et al. (118)	Consulters	Active trunk exercise programme	Chiropractic manual therapy	3	Specified
Cecchi et al. (119)	Consulters	Back School	Manual Therapy	5	Specified
Chown et al. (121)	Consulters	Physiotherapy group exercise	Osteopathy	3	None
Costa et al. (122)	Consulters	Motor control	Placebo	6	Specified
Diaz Arribas et al. (123)	Consulters	Godelieve Denys-Struyf Method	Physiotherapy	3	Specified
Ferreira et al. (124)	Consulters	General exercise; Motor Control Exercise	Manual Therapy	4	Specified
Goldby et al. (126)	Consulters	Spinal Stabilisation	Minimal intervention (Booklet)	6	Specified
Groessl et al. (127)	Consulters	Yoga	Waitlist	8	Specified
Hall et al. (129)	Non-consulters	Tai Chi	Waitlist	5	Specified
Hansen et al. (128)	Non-consulters	Strengthening	Placebo	2	Inferred
Harris et al. (130)	Non-consulters	Strengthening	Brief intervention	6	Specified
Jans et al. (132)	Non-consulters	Cesar Therapy	No further	4	None
Johnson et al. (134)	Consulters	Exercises	GP care	3	Inferred
Russell et al. (138)	Consulters	Exercise Class	Manual Therapy	5	Specified
Saper et al. (139)	Non-consulters	Yoga	Brief intervention (booklet)	6	Specified
Shirado et al. (140)	Consulters	Strengthening and stretching	NSAIDs	4	Specified
Storro et al. (141)	Consulters	Strengthening and stretching	Usual Care	1	Specified
Tilbrook et al. (142)	Consulters	Yoga	Usual Care (booklet)	10	Specified

3.3.7 SMDs of RCTs: Individual and Grouped

SMDs were calculated for each RCT according to the methods described previously in section 3.2.6.1, and the results are summarised in Table 3-7.

3.3.7.1 **Calculation of SMDs per trial**

The SMDs between exercise interventions and control/comparison interventions in the 27 included RCTs ranged from a negative between-arm difference (in favour of the control arm) (-0.96) to very large difference (6.50) in favour of the exercise arm, as seen in Table 3-7. The two outliers with very large SMDs (120,133) were queried as to their accuracy. Chen et al. (120) was queried regarding the accuracy of the reported SDs (pre- and post-intervention) as the control group were very different to the other time-points, and to the intervention group SDs described in the RCT paper. Given this very unusual and unexpected result, the SD was replaced with the SD for the intervention arm at baseline. The authors were contacted to clarify this anomaly but did not respond, thus, this SMD for this RCT was therefore excluded from the meta-analysis (and the other primary outcome was used). Similarly, the very large effect size seen in the Járomi et al. (133) RCT was queried. The authors were similarly contacted for clarification of their reported results and responded that their figures were correct, and this was retained.

Eighteen comparison arms were included in the seven ‘matched’ RCTs: one RCT had three arms, and five RCTs (of seven) used two primary outcome domains.

Fifteen (of eighteen, 83%) SMD results were statistically significant in favour of the exercise arm when compared to a control arm as seen in Table 3-7.

In the twenty RCTs comprising the 'unmatched' group, 35 comparison arms were included. Two RCTs used three primary outcome domains, and four RCTs used two primary outcome domains. Two RCTs had three comparison arms, and one RCT had two comparison arms. Thirteen comparisons (37%) found statistically significant results in favour of the exercise arm in comparison to a control arm, and seven comparison arms (20%) found statistically significant results in favour of the non-exercise control arm in comparison to exercise as seen in Table 3-7.

SMDs using RoMs and follow-up SDs were compared to SMDs using baseline SDs to ensure consistency in the reported figures. Please see Appendix 9.f.ii.Sensitivity Analysis Using Follow-up Standard Deviations

Table 3-7: SMDs in primary outcomes between exercise and comparison arms per RCT

	Trial	**Outcome Domain**	**Comparison Group(s)**	**SMD (95% confidence intervals)**	**Interpretation (as per Cohen** (96)**)**
Matched	Chen et al. (120)	*Pain*	*Ex vs CG*	*0.82 (0.47, 1.17)*	*Large in favour of exercise*
		Exercise Self-efficacy	*Ex vs CG*	*5.01 (4.53, 5.48)*	*Very large in favour of exercise*
	Hildebrandt et al. (131)	*Recovery*	*Ex vs CG*	*0.69 (0.37, 1.0)*	*Medium-large in favour of exercise*
	Járomi et al. (133)	*Pain*	*Ex vs CG*	*6.50 (6.16, 6.83)*	*Very large in favour of exercise*
		Lifting Capacity	Ex vs CG	0.30 (-0.05, 0.64)	Medium in favour of exercise
	Maul et al. (135)	Lifting Capacity	Ex vs CG	0.37 (-0.02, 0.76)	Small- medium in favour of exercise
	Miyamoto et al. (136)	*Pain*	*Ex1 vs WL*	*0.84 (0.62,1.06)*	*Large in favour of exercise*
			Ex2 vs WL	*0.98 (0.80,1.16))*	*Large in favour of exercise*
			Ex3 vs WL	*1.30 (1.07,1.52)*	*Large in favour of exercise*
			Combined vs WL	*1.02 (0.77,1.27)*	*Large in favour of exercise*
		Physical Function	*Ex1 vs WL*	*0.66 (0.58,0.74)*	*Medium-large in favour of exercise*
			Ex2 vs CG	*0.87 (0.79, 0.95)*	*Large in favour of exercise*
			Ex3 vs WL	*1.02 (0.94, 1.10)*	*Medium-large in favour of exercise*
			Combined vs WL	*0.90 (0.49,1.30)*	*Large in favour of exercise*
	Moffett et al. (137)	*Activity Avoidance*	*Ex vs UC*	*0.14 (0.06, 0.22)*	*Small in favour of exercise*
		Physical Function	*Ex vs UC*	*0.41 (0.34, 0.48)*	*Small-medium in favour of exercise*
	Garcia et al. (125)	*Pain*	*Ex vs CG*	*0.49 (0.16, 0.81)*	*Medium in favour of exercise*
		Physical Function	Ex vs CG	0.32 (-0.01,0.66)	Small in favour of exercise
Unmatched	Albaladejo et al. (116)	*Physical Function*	*Ex vs Combined CG*	*0.24 (0.01,0.47)*	*Small in favour of exercise*
	Bronfort et al. (117)	Pain	Ex vs SMT	0.21 (-0.07, 0.50)	Small in favour of exercise
	Cambron et al. (118)	*Pain*	*Ex vs SMT*	*0.38 (0.13, 0.64)*	*Small- medium in favour of exercise*
		Physical Function	Ex vs SMT	0.11 (-0.14, 0.37)	Small in favour of exercise
	Cecchi et al. (119)	Physical Function	*GE vs SMT*	*-0.96 (-1.09, -0.82)*	***Large in favour of SMT***
			IPT vs SMT	*-0.75 (-0.87, -0.62)*	***Large in favour of SMT***

	Chown et al. (121)	Physical Function	Ex vs SMT	-0.00 (-0.52, 0.51)	No effect
	Costa et al. (122)	*Pain*	*Ex vs CG*	*0.49 (0.07,0.90)*	*Medium in favour of exercise*
		Physical Function	*Ex vs CG*	*0.63 (0.20, 1.05)*	*Medium–large in favour of exercise*
		Effect	*Ex vs CG*	*0.53 (0.12, 0.94)*	*Medium in favour of exercise*
	Díaz Arribas et al. (123)	*Pain*	*Ex vs UC*	*1.21 (0.86, 1.56)*	*Large in favour of exercise*
	Ferreira et al. (124)	*Physical Function*	*GE vs SMT*	*-0.70 (-1.19, -0.22)*	***Medium- large in favour of SMT***
			MCE vs SMT	0.05 (-0.44, 0.53)	Small in favour of MCE
			Combined Int vs SMT	***-0.33 (-0.61, -0.06)***	***Small in favour of SMT***
		Effect	*GE vs SMT*	***-0.67 (-1.05, -0.29)***	***Medium- large in favour of SMT***
			MCE vs SMT	0.23 (-0.07, 0.53)	Small in favour of MCE
			Combined Int vs SMT	0.19 (-0.35,0.72)	Small in favour of exercise
	Goldby et al. (126)	Pain	Ex vs CG	0.24 (-0.07, 0.54)	Small-medium in favour of exercise
	Groessl et al. (127)	Physical Function	Ex vs WL	0.14 (-0.18, 0.46)	Small in favour of exercise
	Hall et al. (129)	*Pain*	*Ex vs WL*	*0.52 (0.21, 0.83)*	*Medium in favour of exercise*
	Hansen et al. (128)	Pain	Ex vs CG	0.18 (-0.95, 0.02)	Small in favour of exercise
	Harris et al. (130)	Sick leave	Ex vs UC	-0.16 (-0.48,0.16)	**Small in favour of control**
	Jans et al. (132)	Recovery	Ex vs UC	0.30 (-0.06,0.66)	Small-medium in favour of exercise
	Johnson et al. (134)	*Pain*	*Ex vs CG*	*0.30 (0.04, 0.57)*	*Small in favour of exercise*
		Physical Function	Ex vs CG	0.15 (-0.41, 0.11)	Small in favour of exercise
	Russell et al. (138)	Physical Function	*Ex vs CG*	*0.33 (0.28,0.39)*	*Small-medium in favour of exercise*
			Ex vs SMT	***-0.09 (-0.04, -0.15)***	***Small in favour of SMT***
			Ex vs Combined CG	*0.11 (0.07, 0.15)*	*Very small in favour of exercise*
	Saper et al. (139)	Pain	Ex vs CG	0.21 (-0.09, 0.50)	Small in favour of exercise
		Physical Function	Ex vs CG	0.25 (-0.05, 0.54)	Small in favour of exercise
	Shirado et al. (140)	Physical Function	Ex vs CG	0.27 (-0.02, 0.55)	Small-medium in favour of exercise
		Pain	Ex vs CG	0.18 (-0.12,0.47)	Small in favour of exercise
Unmatc hed		*HRQoL*	*Ex vs CG*	*0.29 (0.00, 0.57)*	*Small in favour of exercise*
	Storrø et al. (141)	*Sick leave*	*Ex vs CG*	*0.74 (0.55, 0.93)*	*Medium-large in favour of exercise*

	Tilbrook et al. (142)	*Physical Function*	*Ex vs WL*	*0.50 (0.26, 0.74)*	*Medium in favour of exercise*

Interpretations highlighted in bold favour the control group; results in italic are statistically significant. The outliers are shaded in grey. Abbreviations used: Ex is exercise, CG is control group, WL is waiting list, SMT is manual therapy, UC is usual care, MCE is motor control exercise, GE is general exercise, Int is intervention.

3.3.7.2 **Calculation of Treatment Success per RCT**

Where possible, treatment success was determined using the recommendation from Dent and Raftery (108) as detailed in section 3.2.6.2. It was not possible to determine a standardised MCID where RCT reports did not include the expected SD and MCID required for their sample size calculation. Only ten RCTs included this information, and of these, only two provided this for more than one specified primary outcome domain (122,136) as seen in Table 3-8.

Of the 22 comparison arms evaluated, only five comparison arms (136) found clinically and statistically significant results in favour of the exercise arm in comparison to a control arm.

Table 3-8: Summary table of SMCID and interpretation per RCT

	Trial	Primary Outcome Domain	Comparison Arms	SMCID	Interpretation
Matched	Miyamoto et al. (136)	*Pain*	*Ex1 vs WL*	*0.54*	*Clinically and statistically significant in favour of exercise*
			Ex2 vs WL		
			Ex3 vs WL		
			Combined vs WL		
		Physical Function	*Ex1 vs WL*	*0.82*	*Clinically inconclusive in favour of exercise*
			Ex2 vs WL	*0.82*	*Statistically significant in favour of exercise*
			Ex3 vs WL	*0.82*	*Clinically and statistically significant in favour of exercise*
			Combined vs WL	*0.82*	*Statistically significant in favour of exercise*
	Moffett et al. (137)	*Activity Avoidance*	*Ex vs UC*	*0.35*	*Inconclusive in favour of exercise*
	Garcia et al. (125)	*Pain*	*Ex vs CG*	*0.54*	*Inconclusive in favour of exercise*
		Physical Function	Ex vs CG	0.82	Inconclusive in favour of exercise
Unmatched	Díaz Arribas et al. (123)	*Pain*	*Ex vs UC*	*0.60*	*Clinically inconclusive and statistically significant in favour of exercise*
	Costa et al. (122)	*Pain*	*Ex vs CG*	*0.50*	*Inconclusive in favour of exercise*
		Effect	*Ex vs CG*	*0.59*	*Inconclusive in favour of exercise*
		Physical Function	*Ex vs CG*	*0.56*	*Inconclusive in favour of exercise*
	Hall et al. (129)	Pain	*Ex vs WL*	0.75	Truly inconclusive
	Albaladejo et al. (116)	*Physical Function*	*Ex vs Combined CG*	*0.75*	*Truly inconclusive*
	Johnson et al. (134)	Physical Function	Ex vs CG	0.50	Truly inconclusive
	Tilbrook et al. (142)	Physical Function	Ex vs WL	*0.39*	*Statistically significant in favour of exercise*
	Russell et al. (138)	Physical Function	*Ex vs CG*	*0.63*	*Statistically significant in favour of exercise*
			Ex vs SMT	**0.63**	**Inconclusive in favour of SMT**
			Ex vs combined CG	*0.63*	*Inconclusive in favour of exercise*

Italics represent results that were statistically significant in favour of exercise arm, Bold represents results in favour of the control arm, and shaded grey cells represent clinically and statistically significant results according to the standardised minimum clinically important difference (SMCID).

3.3.8 Meta-Analysis: Calculation of Grouped SMDs

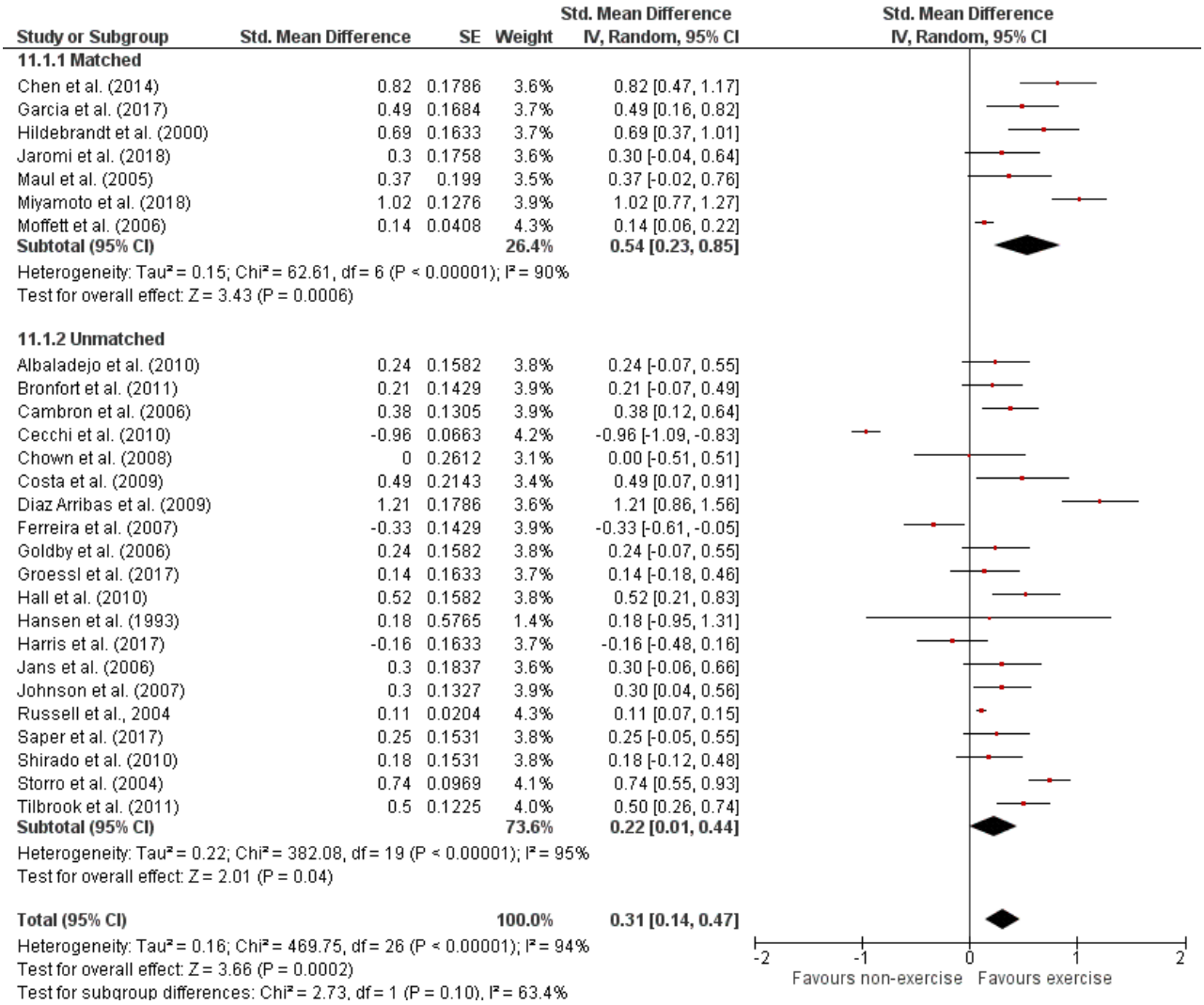
3.3.8.1 **Combined SMD**

Figure 3-8 summarises the overall SMD for all included RCTs and shows that the combined SMD favoured exercise with a small to medium effect size (0.31 (95% CI 0.14; 0.47)) compared to any controls. This result was statistically significant (p=0.0002); however, clinical importance was difficult to judge due to the multiple scales used. Heterogeneity was very high (I^2 = 94%).

3.3.8.2 **Categorised SMDs**

When SMDs were compared across the trials according to categorised status, a medium effect (Cohen, 1992) was seen in the matched trials (SMD 0.54 (95% CI 0.23 to 0.85)) which was statistically significant compared to non-exercise controls in those trials (p=0.0006) as seen above in **Error! Reference source not found.**0. The trials judged to not match their primary outcome to the treatment targets had SMDs of small effect in favour of exercise (SMD 0.22 (95% CI 0.01, 0.44), p=0.04) as seen in **Error! Reference source not found.**0. Total sample sizes in the included RCTs varied from 1197 in the matched category, to 4510 in the unmatched group. The I^2 was consistently high across the matched and unmatched categories with 90% in the matched, and 95% in the unmatched category. This demonstrates that the heterogeneity was high across both groups.

Figure 3-8: Forest plot of SMDs of primary outcome domains, including 'matched' and 'unmatched' RCT sub-groups



SE is standard error; IV is inverse variance; CI is confidence interval; Std. is standard as part of SMD.

3.3.8.2.1 *Sensitivity Analyses*

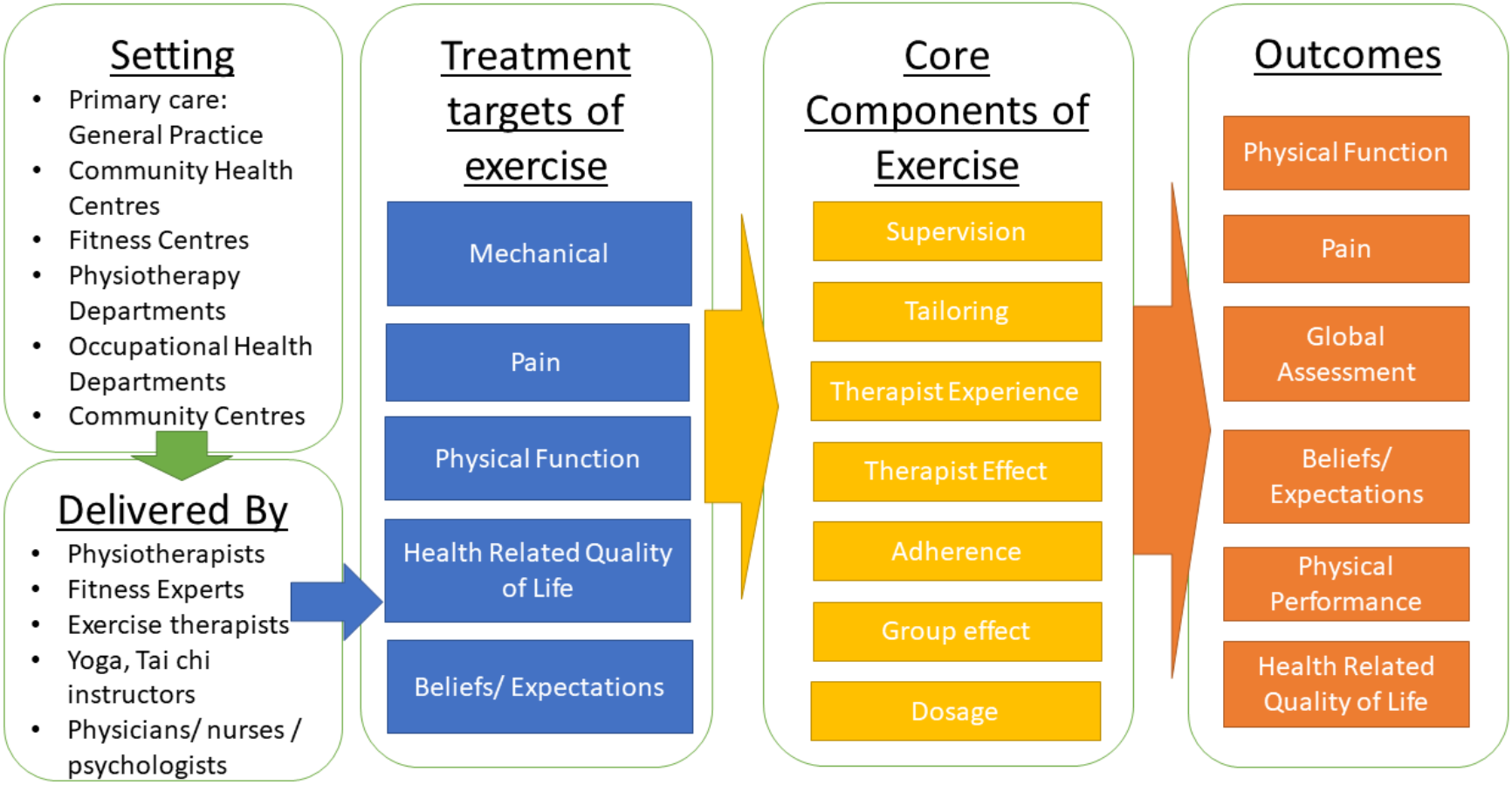
Sub-group and sensitivity analyses demonstrated similar trends, with the 'matched' group of RCTs continuing to demonstrate a larger between-arm effect size with RoM and weighted mean differences (WMD) across pain and physical function outcomes, although the differences between 'matched' and 'unmatched' groups were not statistically significant. SMDs were also compared across participant recruitment strategy, specified (in contrast to inferred) treatment targets, risk of bias (low or high) and comparator groups, within 'matched' and 'unmatched' categories. Similar trends were demonstrated, with the 'matched' group of RCTs yielding a larger SMD than the 'unmatched' group, although, unsurprisingly, none of these sub-group differences were statistically significant (please see Appendix 9.f.iv Effectiveness Sensitivity Analysis and Appendix 9.f.v Sub-group Analyses for a full description of sensitivity and sub-group analyses performed).

3.3.9 Logic Model

As recommended for systematic reviews of complex interventions (13,114), the results of the treatment target and outcome domain extraction were used to develop a preliminary logic model for exercise interventions for persistent NSLBP, summarised in Figure 3-9. The logic model reads from left to right, including the treatment targets, core components of the exercise interventions using data extracted from the included trials, and primary outcome domains. The first column includes the setting and deliverers of the exercise intervention. The second column (treatment targets) was developed by synthesising each included

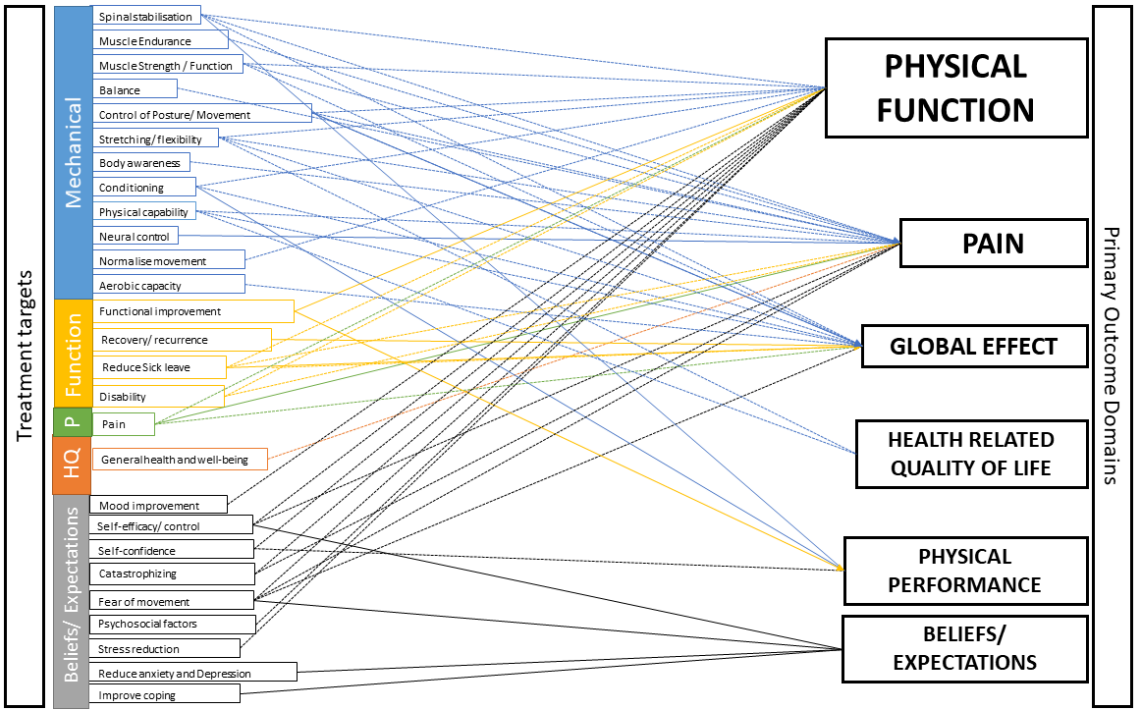
trial's treatment targets. The third column included core components of the exercise interventions. The final column (primary outcome domains) was developed by synthesising the reported primary outcome domains.

Figure 3-9: Logic model depicting reported exercise treatment targets, core components and outcome domains



The relationship between each variable was further mapped for greater detail in Figure 3-10, demonstrating which primary outcome domains mapped onto which stated exercise treatment targets in the included RCTs. For example, the primary outcome domains of pain and physical function were both matched to the treatment target of increasing spinal stabilisation. Similarly, improving muscle strength or muscle function was matched in some RCTs to the primary outcome domains of pain, disability and work status.

Figure 3-10: Figure demonstrating the relationship between exercise treatment targets and primary outcome domains



On the left, the different treatment targets have been categorised by colour according to the type of construct (mechanical constructs in blue, function in yellow, pain in green, heath related quality of life in orange, beliefs/ expectations in grey). On the right, primary outcome domains are listed, with increasing font size depicting increased use. Where dashed lines represent unmatched relationships and solid lines represent matched relationships; P is pain; HQ is HRQoL

3.3.10 Summary of Results

In total, 27 RCTs were included in this systematic review, of which only seven RCTs were categorised as having matched their outcomes to their stated exercise treatment targets. There was a wide variety in the reported treatment targets of exercise in this set of 27 trials, but most treatment targets were specified in only a single RCT, highlighting the lack of consensus about exercise treatment targets in the management of patients with persistent NSLBP. The overall between-arm SMD observed in the RCTs within the ‘matched’ category was larger, in favour of exercise, than the ‘unmatched’ category. Whilst differences were not statistically significant, the trend was the same across all sub-group analyses, suggesting that better matching of the outcomes of RCTs to exercise treatment targets may be more likely to achieve better outcomes for patients with persistent NSLBP. The logic model demonstrated a visual representation of the relationship between the identified primary outcome domains and exercise treatment targets, showing in visual form the poor matching between these in RCTs to date.

3.4 Discussion

This review systematically identified, synthesised and analysed the treatment targets and outcomes of exercise RCTs with patients with persistent NSLBP. It included only those RCTs with adequate statistical power to detect at least medium effect sizes between exercise and comparison interventions. The 27 included RCTs reported 31 different exercise treatment targets, 19 of which were only mentioned in a single RCT. A total of six different primary outcome domains were identified. Only seven of the twenty-seven RCTs (26%) matched the primary outcome domain and measure to the treatment targets of the exercise intervention. The majority of included RCTs (n=20) did not match any of the outcomes to the treatment targets of exercise described. A much greater proportion of RCTs (71%) in the 'matched' category reported statistically significant findings in favour of exercise versus their comparator intervention, in contrast to only 20% of 'unmatched' RCTs. Although meta-analysis found a SMD of medium size for exercise versus comparison in the matched RCTs (SMD 0.54 (95% CI 0.23 to 0.85) (p=0.0006)), and a smaller SMD in favour of exercise in the 'unmatched' RCTs (SMD 0.22 (95% CI 0.02 to 0.44), (p=0.04)), this difference was not statistically significant (p=0.10). There was high heterogeneity, and despite the attempt to include only higher quality RCTs based on sample size at randomisation, RCTs with a high risk of bias were still included. This exploratory review provides some initial support for the hypothesis that better matching of outcome domains to the treatment targets of exercise interventions in RCTs with patients who have persistent NSLBP may be more likely to lead to statistically and clinically significant results, in favour of exercise compared to controls.

3.4.1 Interpretation of the Results

3.4.1.1 **Objective i and iii: Treatment Targets in Exercise RCTs for Persistent NSLBP**

There is clearly a lack of consensus in the published literature about what exercise interventions are trying to achieve for patients with persistent NSLBP. Although there is much evidence to support the benefits of exercise for physical and mental health (26,27,145), the mechanisms of action of exercise and the specific treatment targets of exercise for patients with persistent NSLBP are still a matter of debate (46,47,146). Exercise has been purported to improve motor control, patterns of movement and muscle activation, strength, endurance, flexibility, range of motion, general fitness, as well as mood and depression in persistent NSLBP (27,147). However, despite these wide and varied targets, few RCTs consistently measure these targets, nor report them. This review found many physical performance exercise targets identified by the authors of the included RCTs (e.g. spinal stabilisation, muscle strengthening, stretching and flexibility), which was anticipated given the targeted physical performance aspects of many exercise trials (e.g. strength, flexibility) (22,147). Falla and Hodges (147) describe the mechanical features of spinal pain that exercise can potentially alter such as suboptimal posture or alignment, altered patterns of muscle activation, and altered movement strategies resulting in poorly controlled motion or over-compression. However, the finding that 19 of the 31 different treatment targets were mentioned in only one RCT each demonstrates clear uncertainty about the targets of exercise interventions for persistent NSLBP.

Even in trials testing similar exercise regimes (e.g. McKenzie exercise approach), many different treatment targets were mentioned in the RCT papers (125,137).

There is a discrepancy in the use of outcomes (most frequently pain and disability) that are matched to the exercise treatment targets (148). This is further reinforced by the poor correlation that has been observed between physical performance measures and pain and disability outcomes (22,34,35). Lee et al. (148) assert that treatments use indirect pathways (treatment mediators) to create an effect on outcomes such as pain and disability. It remains to be explored whether better identification of these indirect pathways will provide greater overall effects of treatments or potentially more promising conclusions about treatments from RCTs in LBP, by using the identified treatment targets as treatment mediators. There are few formal tests of the treatment mediators of exercise for persistent NSLBP; those that have been tested include pain catastrophising, which was shown in one study to partially mediate the effect of tai chi exercise on pain intensity, bothersomeness and pain-related disability when compared to waitlist controls (149). Pain catastrophising has also been shown in a further study to mediate the reduction of disability, pain intensity, main complaints and depression in active exercise when compared to no treatment (150). In a knee osteoarthritis population, self-efficacy and pain have been shown to be partial mediators of the beneficial effect of exercise and dietary weight loss on stair climb time in overweight older adults in one study (151). Exploration of treatment mechanisms in other psychological interventions in NSLBP have identified that pain and physical function outcomes are improved by reduced

distress, fear, pain catastrophising and increased self-efficacy (152). Recent work suggests that irrespective of the psychological intervention delivered, the mechanisms underpinning the intervention remain the same (153,154). However, these mediators explain only 20-33% of the relationship between pain and disability (148) suggesting there are other, as yet unidentified mediators (and treatment targets) in the development of pain-related disability in persistent NSLBP.

The mediators that have been tested in studies of exercise for NSLBP are not commonly identified as treatment targets (pain catastrophising was identified as a treatment target in only one RCT(136)) and are even less commonly identified as primary outcomes. The work of Lee et al. (148) and Whittle et al. (44) provides support for the use of more appropriate primary outcomes that correspond to the treatment targets of the intervention used, whilst recognising the importance of patient-specific outcomes. The current systematic review adds exploratory evidence to the argument that better matching of the primary outcome domains to the hypothesised treatment targets of exercise interventions may lead to greater effect sizes in favour of exercise and therefore ultimately change the conclusions of exercise RCTs.

3.4.1.2 **Objective i and iii: Outcome Domains used in Exercise RCTs for Persistent NSLBP**

It is perhaps unsurprising that the most frequently cited primary and secondary outcome domains found in this review were those recommended in core outcome

sets for RCTs in the field of LBP (pain and physical function) (74,78,155). Thirteen included RCTs measured all three of the core domains (pain, physical function and HRQoL) (112-116,121,125,126,131,133,136,137,139), and for all three of these domains, the most frequently used outcome measures were in line with published recommendations (82). Gianola et al. (68) in their systematic review of 185 trials of rehabilitation interventions in LBP similarly reported that pain was measured in 89% of trials, and physical functioning in approximately 64% of trials. In the Delphi study performed by Chiarotto et al. (82) patients and clinicians placed particularly high emphasis on psychological functioning, and this has been reflected in the included trials in this review that used psychological outcomes as the fourth most frequently cited secondary outcome domain. A wide variety of psychological constructs and measures were identified in this review, in contrast to that reported by Chapman et al. (86) in which fear-avoidance, depression and anxiety were the most commonly cited psychological outcomes in LBP research. In this review, the psychological treatment targets that were most frequently reported were self-confidence (135,138,142) and fear of movement (124,130,136), followed by stress reduction (127,139) and self-efficacy (120,138).

3.4.1.3 **Objective ii: Risk of Bias and Quality of Reporting of Included RCTs**

This review used eligibility criteria that focused on including trials with adequate statistical power to detect at least a medium difference between intervention-arms, and thus (at least in theory) trials of higher quality (97). Despite the

restrictive eligibility criteria, only six trials included in this review met one domain of high risk of bias, and only half of the included trials met at least three of the seven domains of the risk of bias criteria (14/27). In support of these criteria, Rubinstein et al. (97) found in their post hoc analysis that spinal manipulation trials in LBP that had performed a sample size calculation were more likely to have lower risks of bias than those who did not perform a sample size calculation. Whereas Froud et al. (53) in their review of sample size calculations in trials of LBP found that only one-third of trials were powered to detect a SMD of less than 0.5 and only 5% were powered to detect SMDs of less than 0.3. Hence, the majority of trials in this review found small SMDs in favour of exercise.

The recent release of the Cochrane ROB 2.0 provides more detailed guidance for assessing both trial per-protocol-analysis as well as their intention-to-treat analysis and provides algorithms for reaching a decision on their risk of bias (156). This guidance would reclassify many of the trials that were classified as high risk of bias in the domain of blinding of participants/ personnel, as this guidance establishes whether allocation concealment generated any deviations from the intended intervention delivery in the trial. If not, then the trial is recommended to be categorised as low risk of bias in this domain. This would have resulted a total of twelve of the twenty-seven trials being categorised as low risk of bias (44%).

The Consolidated Standards of Reporting Trials (CONSORT) statement (63) checklist was designed to improve the reporting and replicability of RCTs

(64,157). Alongside this, the Template for Intervention Description and Replication (TIDieR) checklist aimed to improve the description of interventions included in RCTs. Reporting of both treatment targets and primary outcome domains appears to have improved since the publication of the CONSORT (63) and TIDieR (64) statements but, it is difficult to draw conclusions based on the limited trial data available from this review. In this review, of the eight trials that did not specify exercise treatment targets: two were published after 2010, one in 1993, and the other five between 2000 and 2008. In contrast, all six trials published after the publication of the TIDieR guidance (64) specified their treatment targets. Point 6a of the CONSORT statement details “pre-specified primary and secondary outcome measures”: in this review, four of the five trials that did not specify primary outcome measures were published prior to 2010. Candy et al. (158) reviewed pre- and post-publication RCTs against the CONSORT complex interventions extension (159) and found similar rates of reporting, with most trials providing adequate rationale for their intervention: 93% pre-2002-2007 and 99% post-2010-2015. However, they found no evidence that the overall standards of reporting of the intervention content in complex interventions were improving over time. Glasziou et al. (160) and Schroter, Glasziou and Heneghan (161) reviewed the reporting of interventions in trials published in the British Medical Journal and found that 57% of trials included in their review did not provide adequate descriptions of the interventions, with 87% of the interventions in back pain trials judged to be non-reproducible (161). They highlighted the need for improved descriptions of interventions, in particular the sequencing of the intervention, the materials used, the dosage and duration of

delivery, and the schedule of delivery. In exercise interventions, these have been addressed within the Consensus on Exercise Reporting Template (CERT) (65), which has been developed to address these concerns. However, the CERT does not specify the need to identify the treatment target(s) of the exercise intervention, which is a shortcoming identified by Kent et al. (66) and Wood, Ogilvie and Hayden (67).

3.4.1.4 Objective iv: the SMDs of Exercise in Comparison to a Control Treatment

The overall SMD from the included RCTs was 0.31 (classified as small in favour of exercise versus controls), which is similar to other systematic reviews of exercise for LBP (26,27,29). This demonstrates that the exclusion of small trials did not, on the whole, change the result regarding the average effects of exercise compared to other interventions. The findings of this systematic review are, therefore, likely to have been similar had smaller trials been included (n=29 excluded on sample size), as smaller trials would have had less power to detect SMDs and would likely have had higher risks of bias (97). Searle et al. (27) found similar SMDs to these results when comparing pain outcomes of exercise to a control or other treatment group (SMD -0.32 (95% CI -0.44,-0.19), ($p<0.01$)) (in favour of exercise); which is comparable to the total SMD of 0.31 ((95% CI 0.14; 0.47) ($p=0.0002$)).

SMDs were used in this review as recommended by the Cochrane Handbook 5.1 (106); however, there is some debate about whether this is recommended for

combining different scales measuring the same construct (162). In response to these concerns, RoM analysis was also performed (110), as well as pooled WMD scores for trials reporting pain and physical function scores. SMDs have been found to be more generalizable than the pooled WMD (163). Although SMD has the potential for over-or under-estimation depending on the variability of the patients (due to the use of the standard deviation), Takeshima et al. (163) propose that the SMD has greater generalisability than the use of mean difference. In this review, one trial reported results with potentially incorrect standard deviations for the control group (120), resulting in the possible overestimation of the SMD and leading to the exclusion of this RCT from further analysis. The sensitivity analysis performed using the RoM replicated this finding, with a medium effect size in the matched category (22% in favour of exercise in comparison to a non-exercise control), and a small effect size in the unmatched category (7% in favour of exercise in comparison to a non-exercise control). This suggests that using the SMD or the RoM produced similar findings in this review.

3.4.1.5 **Objective v: Process-Orientated Logic Model**

A preliminary logic model was developed using the data from the RCTs included in this review to assist with interpretation and understanding of the relationships between the treatment targets of exercise and the outcomes used in trials of exercise interventions. This helped to visualise these relationships. However, due to the wide variety of constructs within both the treatment targets and outcomes, it was challenging to include them all in a clear logic model. Rohwer et al. (13) highlight that the use of a logic model "does not have to be a perfect

reflection of the world but should depict the assumptions contained in the review". Further, they recognise the risk that depicting a complex system within a logic model may lead to overcrowding of the logic model with information, with the result that it is difficult to interpret the graphic without considerable accompanying text (13). The logic model presented in Figure 3-9 has potential utility for future teams developing exercise interventions as part of RCTs as it encompasses: the components contributing to the complexity of the exercise intervention, the context (including location and deliverer of the exercise intervention), the treatment targets of the intervention as well as the most appropriate outcomes to be used. This review highlights the need for future trials of exercise interventions to use logic models to underpin their intervention, as the process of developing logic models means that the targets of the intervention need to be made explicit, thus showing how their choice of outcomes clearly relates to the exercise targets. The logic model presented in this text is an example of what factors may be important to consider in future trials of exercise interventions. The use of logic models allows a visual flow from the treatment targets of the exercise, to the exercise selection, dosage and administration, to the most appropriate outcome measures and domains.

3.4.2 Strengths and Limitations of this Review

This review is the first that has explored the potential impact of matching the primary outcomes to reported treatment targets in RCTs of exercise for persistent NSLBP. This review was performed both according to a published protocol as well as the PRISMA guidance, ensuring a rigorous approach (94). Independent

reviewer selection, quality assessment and data extraction were performed by pairs of reviewers, strengthening the reliability of results (101).

The likelihood of publication bias should be considered in light of the results gained from this review, as trials with more favourable results are more likely to have been published (164), and therefore included in this review. Excluding RCTs on sample size is not established practice in systematic reviews, and may have led to the unnecessary exclusion of some smaller, but high-quality RCTs. The nature of this exploratory review was limited by the reliance on using available published data: a further limitation may be that RCTs might have indeed tested exercise interventions that did match their primary outcome(s), but since this was not stated in published papers, it was assumed that they did not. The fidelity of the exercise interventions was not assessed within this review, and it is perhaps possible that exercise interventions which are delivered to high fidelity lead to better patient outcomes. This review included a small sample of heterogeneous exercise interventions, comparator interventions, population characteristics, outcomes used and follow-up periods. This comparison of matched versus unmatched RCTs was, by necessity, a non-randomised comparison, and the analyses were tested for other potential explanations for the results by considering sub-group analysis based on comparator interventions.

The heterogeneity of all meta-analyses performed was high in this systematic review. Pooling RCTs when the heterogeneity is high is expected when

combining trials of different exercise interventions, in different settings (25). However, this was controlled for by using a random-effects model (106). When heterogeneity levels are high as was seen in this meta-analysis, the standard error of the pooled estimate is likely to be very large, which results in low power for the corresponding test to detect a difference between groups (165). The majority of included trials in this review found small between-group SMDs in favour of exercise, and like all RCTs, there may be other factors that account for these effect sizes that are unaccounted for (106). The difference between the 'matched' and 'unmatched' groups of RCTs was not statistically significant, perhaps related to the relatively small number of included RCTs. This review was exploratory in nature, and was likely underpowered to detect a statistical difference between the 'matched' and 'unmatched' groups (165). This limits the interpretation of the results due to a lack of strength of evidence, but it would be advisable to replicate this analysis in a larger group of RCTs (that was sufficient to test for between-group differences) to see if these findings are replicated. Had sample size not restricted inclusion into this review, more studies may have been included (n=29), which may have been of high quality and low risk of bias, but would have been unlikely to be sufficiently powered to detect a between-group change. The identification of treatment targets has been replicated in the Cochrane review of exercise trials (67)(n=265) with similar results reported to those of this review, which suggests the findings for this review may have been similar had smaller trials been included.

In the calculations for the SMD, average SD were used with average SMD values. This may have impacted the calculations of the SMD when sample sizes varied across arms which was present in 44% (12/27) of trials included. (Please see Appendix 9.f.vi Sample size analysis across arms of included trials).

The categorisation of RCTs into matched categories was a subjective process, and should this have been performed by a different team, different conclusions may have been drawn. Although each RCT was independently judged by two researchers (from a team of four), it was not a formally validated process as none exists to our knowledge. The use of logic models in this chapter demonstrates a clear visual model of the results of data extraction, developed by the PhD candidate. Although, these findings will be strengthened by input from multiple stakeholders, the logic model is a preliminary draft that may have value when considering further trials of exercise intervention for persistent NSLBP.

A strength of this review was the replication of results within all the sensitivity analyses performed. To improve interpretability, WMD of pain and physical function scores was also performed. However, these values were interpreted in line with the recommendations of Ostelo et al. (113), which were originally designed for patient-level improvement not population-level interpretation.

3.4.3 Implications for Clinical Practice and Research

The findings suggest that RCTs that better match their outcomes to their exercise treatment targets may, on average, show greater SMDs from exercise compared to controls than those that do not, and may be more likely to find a statistically significant result in favour of the exercise intervention. This means that the effectiveness of exercise interventions for LBP may have been previously underestimated, because many trials use primary outcomes that are not reflective of the targets of their interventions.

This review identified little consensus about the treatment targets of exercise for NSLBP. This may impact the appropriate selection of both exercise programmes in patient management as well as the outcomes with which to try to assess the effectiveness of exercise programmes. When treating patients with persistent NSLBP, the results of this review suggest that if one can use the most appropriate outcomes to best reflect the treatment targets of exercise, patients may demonstrate greater benefits which will support the use of exercise therapy in rehabilitation. Clinicians need to know what an exercise intervention is targeting and what the effects are of that targeted intervention to determine whether it is worth applying clinically.

The wide variety of treatment targets included in the RCTs in this review emphasises a lack of consensus. This lack of consensus about exercise treatment targets in persistent NSLBP needs to be addressed to help ensure future RCTs can better target the components felt to be most important, and

select the most appropriate outcomes with which to compare clinical effectiveness. Mediation analysis may be a useful way to better understand whether treatment targets contribute to the mechanisms of action identified, but these require RCTs designed to permit this analysis with a range of possible targets and/or mediators identified *a priori*. This should ideally be reflected in a logic model so that the treatment targets are clearly specified and their key outcomes selected to match those treatment targets. Further analysis is also required to understand whether one single outcome, multiple outcomes or a composite outcome (85,166) is most appropriate or useful in trials of complex interventions, such as exercise in persistent NSLBP.

3.5 Conclusion

The primary aim of this systematic review was to investigate the role of matching between reported outcome domains and exercise treatment targets. This review found that matching the primary outcome to the identified exercise treatment targets tends to result in greater overall differences between exercise programmes and non-exercise controls. Whilst this difference was not statistically significant, the same trend was found in 7 sensitivity analyses. 71% of matched trials found statistically significant results in favour of exercise, in comparison to 20% of unmatched RCTs. Most RCTs did not match their primary outcome to the treatment targets of the exercise intervention. This review identified a lack of consensus about the treatment targets of exercise for patients with persistent NSLBP, and improved understanding of these may facilitate clinical practice and future research (see Chapter 6). A ‘partially matched’ group of RCTs that

included some outcomes that reflected their exercise treatment targets, albeit not as their primary outcome, was identified. Further data analysis is required to assess whether using a matched outcome might change the conclusion of these RCTs. The next chapter describes secondary data analysis of the partially matched RCT datasets to compare the results when using a matched outcome versus an unmatched outcome.

4 Chapter 4: Exploratory Secondary Data Analysis of RCT Datasets

Summary

This chapter describes the secondary analyses of five RCTs. In the datasets, the matched secondary outcomes were reanalysed in the same method as the primary outcome. The analyses compare the results of the matched outcomes (in SMD) with the original primary (unmatched) outcome, and explore whether the results and conclusions of the RCTs may have been different as a result.

The chapter and results have been written up as part of a paper:

Wood L, Foster NE, Lewis M, Bronfort G, Groessl E, Hewitt C, Miyamoto G, Reme SE, Bishop A. Matching the outcomes to treatment targets of exercise for low back pain: does it make a difference? Results of secondary analyses from individual patient data of randomised controlled trials and pooling of results across trials in comparative meta-analyses. Under review.

4.1 Introduction

The systematic review in the previous chapter (chapter 3) found that most previous RCTs of exercise in persistent NSLBP did not match their primary outcome to their own identified treatment targets. However, the meta-analysis results suggest that in RCTs that did match their primary outcome to their stated treatment targets, the SMD of the primary outcome was larger than those in RCTs that were unmatched, although this difference was not statistically significant. In

this chapter, this is further explored by undertaking secondary data analyses of several previous RCTs. The identified RCTs included those wherein the primary outcome did not match the stated exercise treatment targets, but (some) secondary matched outcomes were included that captured (some of) the identified treatment targets. It is unknown whether using an outcome matched to the treatment targets might alter the findings of these included RCTs. This chapter presents the findings from secondary analyses of these identified RCTs to explore whether using a matched outcome in place of an unmatched primary outcome might alter the results and conclusions of these RCTs.

4.1.1 Study Aim and Objectives

Aim: This secondary analysis aimed to explore the impact of matching outcomes to the treatment targets of exercise in persistent NSLBP trials on the results and conclusions of RCTs.

Objectives:

i. To perform the analysis applied to the primary outcome(s) by the authors of the RCTs on the included secondary outcome(s) that match their stated exercise treatment targets.
ii. To compare the results of the calculated standardised mean differences (SMDs) using matched secondary outcomes with the SMDs of the nominated primary outcomes of RCTs.

4.2 Methods

4.2.1 Sample

For this exploratory secondary analysis, the five RCTs that formed the 'partially matched' category classified in the systematic review (section 3.3.6 in chapter 3) formed the sample. These RCTs were included, as some or all of the stated exercise treatment targets were captured in the set of secondary outcomes, but the primary analysis performed by the RCT authors was based on an unmatched primary outcome. For example, in one RCT the authors specified treatment targets of improving strength and endurance, the primary outcome was pain, but included in secondary outcomes were measures of strength and endurance (117).

.

4.2.2 Data Extraction

Pertinent information to inform this data analysis had already been extracted by the PhD candidate as part of the systematic review (chapter 3) as a part of independent pairs of reviewers, as follows:

i. The PhD candidate extracted the specified treatment targets of the exercise intervention within each RCT report
ii. The primary and secondary outcome(s) measured and reported for each RCT
iii. The outcomes that matched the stated exercise treatment targets
iv. The PhD candidate extracted the method of analysis performed using the matched secondary and unmatched primary outcomes (e.g. ANOVA, linear mixed model etc.)

4.2.3 Data Analysis

The secondary analyses reported in this chapter varied across the included RCTs and was dependent on whether the RCT dataset was made available for analysis or whether sufficient information was provided in the published RCT paper. Of the five RCTs included in this analysis, initial SMD analyses were performed on three RCT paper results (in one RCT (127), only one secondary outcome was included in the published RCT results). Three datasets were obtained for further analyses. For each included RCT, the information used was dependent on the identified treatment targets and the matched outcome domains extracted in the systematic review (chapter 3), as follows:

i. In RCTs where the analysis performed on the matched secondary outcomes was the same as that performed on the primary outcome measure, and sufficient information was provided by the authors, the SMD was calculated. The resultant SMD was compared to the SMD derived from the primary outcome measure. One RCT (142) provided sufficient information in their trial publication to achieve this. In RCTs where the analysis applied to the primary outcome domains was not repeated on the secondary outcomes by the study authors, or insufficient information was provided for calculation of the SMDs, the RCT dataset was requested from the RCT corresponding author. In total, four datasets were requested (117,127,130,140) and three were obtained (117,127,130).

4.2.3.1 **Standardised Mean Difference (SMD) Calculation**

SMDs were calculated for each primary and matched secondary outcome for between-arm differences at the primary outcome time-point. The primary outcome time-point designated by the RCT authors was used, or the soonest time-point post-exercise-intervention, if no primary time-point was specified by the authors. As in the systematic review (chapter 3), the SMD calculation used the formula:

$$d = \frac{\dot{X}_1 - \dot{X}_2}{s} \qquad 4\text{-}1$$

Where d represents the SMD, Ẋ represents the mean follow-up score, 1 represents the intervention arm, 2 represents the control arm, and s represents the average of the baseline standard deviations (106)*.*

95% CIs were calculated for the SMD, as described in section 3.2.6.1 in chapter 3. SMD statistics for all between-arm differences were given based on intervention minus control: positive SMDs indicate higher values for the exercise intervention (lower for the control), and by contrast, negative SMDs indicate lower values for the intervention (higher for the control). Since the direction of scale scoring may vary, i.e. higher values indicate worse health outcome status (for some scales) and better health status (for other scales) – for purposes of standardisation and ease of evaluation and interpretation within the meta-analysis all SMDs were scaled such that positive SMDs reflected better outcome for the exercise intervention arm and negative SMDs reflected worse outcome for the exercise intervention arm by multiplication with minus one (106).

4.2.3.2 **Secondary Analyses Conducted on Obtained Datasets**

This exploratory analysis was conducted on three obtained datasets. All analyses used IBM Statistical Package for Social Science (SPSS) Statistics 24. The analyses progressed as follows:

4.2.3.2.1 ***Data Modification***

Data in the obtained datasets were modified in a step-wise approach as follows:

i. Datasets were checked for missing data by using descriptive analysis to check the range of inputs per matched outcome and transformed to account for excessive range where necessary (e.g. where missing data was coded with 999 or 768, these values were recoded as system missing).

ii. For linear mixed models the data were transformed from wide to long by transforming the variables to cases and computing a new variable consisting of all time-points relevant to that outcome. For example, the outcome of Pain (new long variable) would include Pain Baseline, Pain 6-week follow-up, 12-week follow-up, and 6-month follow-up. The participant ID would remain the same across these time-points, and other values such as group allocation would also remain the same.

iii. Targeted outcomes were then converted to standardised outcomes in SPSS so that results were comparable across a variety of outcome measurement scales, for example, an outcome “Pain” would be standardised and saved as “ZPain”.

4.2.3.2.2 *Secondary Analyses*

i. Initial analyses aimed to replicate the published or presented data used for the primary outcome(s) and or targeted secondary outcomes where possible to do so. The replicated analysis was applied to the matched secondary outcome(s).

   a. In two datasets, the analyses replicated were ANOVA (130) and ANCOVA (117). These analyses used the primary time-point. Neither of these published RCTs specified which post hoc corrections were used and a variety of corrections were used when trying to replicate the results including Tukey, Scheffé and Bonferroni (167).

   b. In a further two datasets (136,137) (and in one dataset as a second analysis method (117)) linear mixed model analyses were used on the primary outcomes. Linear mixed model analyses include all time-points available for the relevant outcome, and therefore values for all available time-points for the matched secondary outcomes were also used and reported. Models were fitted including patient-identifiers as a random-effect term and including fixed-effects terms in accordance with the trial authors' specification.

ii. Further analyses were performed using the matched outcomes and the method of analysis used for the primary outcome.

4.2.3.3 **Meta-analyses performed on Summary Results**

A summary of the results was produced in forest plots using RevMan (5.3). This was created by including the first mentioned unmatched primary outcome SMD and the first mentioned matched treatment target SMD for each analysed RCT in a sub-group comparison using random effects and the generic inverse variance method.

A further between-group difference and associated 95% CI was calculated by:

i. Generating the difference for each paired between-group difference, and the overall mean between-group difference
ii. Calculating the SD and SE of the mean between-group difference
iii. Calculating the upper and lower CI limits

4.3 Results

4.3.1 Included RCTs

Four datasets were requested, and three were obtained (117,127,130). Shirado et al. (140) were not able to provide a dataset for further analysis due to loss of contact with the dataset holder, and therefore, a limited analysis only was possible using their published data. Tilbrook et al. (142) provided sufficient information in their published report for the SMDs to be compared across primary and secondary outcomes and therefore, their original trial data were not requested. Data sharing agreement for the three trial datasets are included in Appendix 9.g:Data Sharing Agreement for Datasets used in Chapter 4. A summary of stated treatment targets, matched outcome domains and measures, and methods of analysis for each of the included trials is included in Table 4-1.

Table 4-1: Treatment targets, matched outcome domains, and analysis conducted for each included RCT

	Trial	Identified Treatment Targets	Outcome Domains Captured		Primary Time-Point	Analysis Performed	
			Primary	**Secondary**		**Primary Outcome Analysis**	**Secondary Outcome Analysis**
Obtained original datasets	**Bronfort et al.** (117)	Increase trunk muscle endurance and trunk stability	Pain (11-point box scale)	Static endurance (flexion, extension), dynamic endurance (flexion, extension), isometric strength (flexion, extension).	12-weeks*	Analysis of covariance (ANCOVA) to analyse for differences between the three groups. Baseline values were covariates. Linear mixed-model longitudinal analyses (which accounted for correlation over time within participants) using the MIXED procedure in SAS 9.1.	Change scores for trunk performance measures calculated using end treatment (Week 12) and baseline values: analysed for group differences with analysis of variance (ANOVA).
	Groessl et al. (127)	Increased strength and flexibility Stress reduction Increased pain tolerance.	Physical function (RMDQ)	Pain (Brief Pain Inventory) (reported), range of motion (Saunders digital inclinometer) and core strength (prone and supine bridge) (not reported)	12-weeks	Linear mixed-effects modelling to examine the change score across measured time points. A main effect of group (yoga versus delayed treatment), a main effect of time (categorically coded for baseline, 6-weeks, 12-weeks, and 6-months) and an interaction between group X time included in the model.	
	Harris et al. (130)	Fear-avoidance and movement phobia Re-establish normal movement patterns	Increased work participation – change form full-time sick	Fear-avoidance beliefs (Fear-Avoidance Beliefs Questionnaire)	52-weeks	Differences between groups measured with chi-square tests for each of the 12 months.	A mixed between–within-subject analyses of variance with one between-group factor (BI, BI + group CBT,

			leave to partial sick leave or full return to work.				BI + group PE) and with one within-subjects/ repeated measures factor (baseline and 12 months follow-up used. The effect of time and the interaction effect (Time x Group) were reported.
Published Data Only	**Shirado et al.** (140)	Increasing overall physical activity Spinal mobility	Pain (VAS), Physical function (RMDQ) and Health-related quality of life (JLEQ)	Flexibility (finger floor distance)	8-weeks	Unable to replicate primary analysis; only SMD values calculated as per 5.2.3.	
	Tilbrook et al. (142)	Improving mobility Strength Posture Reducing pain	Physical function (RMDQ)	Pain (Aberdeen Back Pain Scale)	12-weeks	Unable to replicate primary analysis; only SMD values calculated as per 5.2.3.	

*Bronfort et al. (115) did not specify their primary time point thus the first time point post-treatment was used, as per the method used in the systematic review. Where RMDQ represents Roland Morris Disability Questionnaire, BPI Brief pain inventory, DT delayed treatment, BI brief intervention, CBT cognitive behavioural therapy, PE physical exercise, SF-12 Short Form 12, PSEQ Pain self-efficacy questionnaire, VAS visual analogue scale, JLEQ Japanese Low Back Evaluation Questionnaire, NSAID non-steroidal anti-inflammatory drugs.

4.3.1.1 **Outcome Measures Used**

A total of five different patient-reported outcome domains and six different outcome measures were used in the five included RCTs. The outcome measures used across trials varied, both in the direction of effect, scale and range. For ease of interpretation, the outcome measures used in the included trials are summarised in Table 4-2.

Table 4-2: The outcome domains and measures used by included RCTs

	Outcome Domain Measured	Outcome Measure/ Tool	Range	Direction of Effect	Interpretation
Patient-reported Outcome Measures	**Return to Work**	Department of Labour reported value (130)	Percentage	←	Greater proportion indicates greater return to work.
	Physical Function	Roland Morris Disability Questionnaire (RMDQ)(127,140,142)	0-24	→	0 “no disability” and 24 “maximum disability”.
		Oswestry Disability Index (ODI) (130)	0-100	→	0 “no disability” and 100 “maximum disability”.
	Pain	Visual Analogue Scale (VAS) (140)	0-10	→	0 “least” to 10 “most intense pain” over the last few days.
		Ordinal 11-point Box scale (117)	0-10	→	Pain over the past week, with 0 “no pain” and 10 “worst pain possible”.
		Aberdeen Back Pain Scale (ABPS) (142)	0-100	→	Higher scores reflect poorer back pain-related function.
		Brief Pain Inventory (BPI) (127)	0-10	→	Pain severity subscale: Four 0-10-point numeric rating scales: rating pain “at its worst” and “at its least in the last 24 hours” and the other two asking about pain “on average” and “right now”. For each NRS, the verbal descriptors are “no pain” and “pain as bad as you can imagine”. Seven other questions relating to the pain interference subscale.
	Fear-Avoidance Beliefs	Fear-avoidance Beliefs Questionnaire (FABQ) (130)	0-112	→	A high score indicates a high level of fear-avoidance. Two subscales – one related to work and another to physical activity.

	Health-Related Quality of Life: Japanese Specific Physical Function and psychosocial concepts	Japan Low back pain Evaluation Questionnaire (JLEQ) (140)	0-120	→	The higher the score, the worse the quality of life.
Objective Measures	**Range of motion**	Saunders digital inclinometer (127)	Measured in degrees	←	Greater range, greater function.
	Trunk Flexibility	Finger-floor distance (140)	Measured in centimetres	→	Greater the size the greater the impairment.
	Core strength	Prone and supine bridge positions (127)	Time maintained in pose to a maximum of 90 seconds	←	Greater time, the better the function.
	Static endurance (flexion, extension)	Biering-Sorensen test (117)	Length of time able to maintain pose	←	Greater time, the better the function.
	Dynamic endurance (flexion, extension)		Maximum number of repetitions	←	Higher number of repetitions, the better the function.
	Isometric strength (flexion, extension)	Computerised digital myograph (DM2000) (117)	Unclear	←	Unclear.

4.3.2 Data Analysis

4.3.2.1 **Standardised Mean Difference Calculations**

SMDs were calculated for the outcomes of interest in three RCTs (127,140,142) as described in section 3.2.6.1, in chapter 3. One of the RCT's primary outcomes were reported using the M-est function, median and inter-quartile (140). These functions were used in contrast to the mean and SD, which relies on the normal distribution of the data, as the data had asymmetrical distribution or were skewed. These SMD results are summarised below in Table 4-3. Bronfort et al. (117) and Harris et al. (130) did not provide sufficient information in the published trial report for SMD calculation of the secondary matched outcomes.

Table 4-3: Calculated SMD values comparing the reported primary outcome domain and matched secondary outcomes

Trial	**Interventions Compared**	**Outcome Domain (Primary outcome shaded)**	**Standardised Mean Difference (95% Confidence Interval)**
Groessl et al. (127)	Yoga vs Waiting list	Physical Function	0.14 (-0.27, 0.55)
		Pain	0.30 (0.08, 0.52)
Shirado et al. (140)	Exercise vs NSAIDS*	Pain	0.17 (-0.12,0.47)
		Physical Function	0.27 (-0.02, 0.55)
		Health-Related Quality of Life	0.29 (-0.00, 0.57)
		Forward Finger Distance	0.54 (0.26, 0.83)
Tilbrook et al. (142)	Yoga vs Usual Care	Physical Function	0.50 (0.26; 0.74)
		Pain	-0.01 (-0.23, 0.22)

*Where NSAIDS refers to non-steroidal anti-inflammatory drugs; Shaded lines represent each trial's stated primary outcome(s) and results.

Of these three RCTs, two displayed greater SMDs that were statistically significant using the matched secondary outcome than using the unmatched primary outcome (117,127). Only Tilbrook et al. (142) demonstrated a larger SMD in the primary unmatched outcome than in the matched secondary outcome.

4.3.2.2 **Secondary Analyses Performed on Datasets Obtained**

4.3.2.2.1 ***Secondary analysis of Bronfort et al.*** (117)

The Bronfort et al. RCT (115) included four time-points (baseline, 12-weeks, 26-weeks, and 52-weeks) and six matched outcomes of interest (dynamic endurance flexion and extension strength, static endurance flexion and extension strength, isometric flexion and extension strength). However, these variables were only measured at baseline and 12-weeks (primary time-point). They included three comparison arms: a specific exercise arm (SET), spinal manual therapy arm (SMT), and a home exercise and advice arm (HEA). The stated primary outcome was pain, and no statistically significant between-arm differences in the primary outcome were reported at any of the time-points. The SMD of each comparison was calculated for the primary time-point using the methods described in chapter 3, section 3.2.6.1 and are tabulated in Table 4-4.

Table 4-4: Published trial results of the primary outcome measure (pain) in Bronfort et al. (2011)

PAIN	**Wk. 12 (Primary)**
SET* vs SMT	0.1 (-0.6,0.8)
SMD	*0.21 (-0.5, 0.07)*
SET* vs HEA	0.6 (-0.1, 1.4)
SMD	*0.43 (-1.03,2.23)*
SMT* vs HEA	0.6 (-0.2, 1.3)
SMD	*0.20 (-1.29,1.89)*

* Where HEA represents home exercise and advice, SMT represents spinal manual therapy, SET represents specific exercise therapy and SMD represents standardised mean difference; Negative values favour the intervention group (denoted with an asterisk, where the intervention is the active comparator, e.g. mean score (SET) – mean score (SMT) for the first data row); italicised data were calculated SMD values.

4.3.2.2.1.1 Data modification

Of the twelve original variables (six matched outcomes at two time-points, baseline and 12-week follow-up), there were no missing data, and all variable

scales scored in the same direction. For the linear mixed model to be run, the dataset was transformed from wide to long, by merging the twelve original variables into six standardised variables as described in section 4.2.3.2.1.

4.3.2.2.1.2 Replicating the Original Results: ANOVA Analyses

It was not possible to repeat the analysis performed on the primary outcome as this was not requested in the dataset (as it was not one of the matched exercise treatment targets). To replicate the published analysis performed on the secondary outcomes, analysis of variance (ANOVA) was used (see Table 4-5). The requested secondary outcomes were analysed with an ANOVA on the change scores (calculated as the difference between the 12-week follow-up and the baseline data). The significance values were reported in the trial paper, and these were replicated using a one-way ANOVA using Tukey's, Scheffé and Bonferroni post hoc analysis (the published paper did not specify which post hoc correction was used). SMDs of change scores are also presented within Table 4-5.

Table 4-5: Replication of ANOVA significance results of the change scores in comparison to the reported results of Bronfort et al. (2011)

Outcome Change Score	**Arm Comparison**	**Post Hoc Analysis: Statistical Significance**			**Standardised Mean Difference (95% CI)**	**Published Significance (Bronfort et al. 2011)**
		Tukey	**Bonferroni**	**Scheffé**		
Static endurance Flexion	SET* vs SMT	<0.0001	<0.00001	<0.00001	0.70 (0.37, 1.03)	<0.0001
	SET* vs HEA	**0.001**	**0.001**	**0.002**	0.53 (0.18 ,0.88)	**<0.0001**
Static endurance Extension	SET* vs SMT	0.011	0.012	0.015	0.42 (0.08, 0.76)	<0.05
	SET* vs HEA	**0.002**	**0.002**	**0.003**	0.53 (0.17, 0.88)	**<0.001**
Dynamic Endurance Flexion	SET* vs SMT	<0.0001	<0.00001	<0.0001	0.75 (0.43, 1.08)	<0.0001
	SET* vs HEA	<0.0001	<0.00001	<0.0001	0.85 (0.51, 1.19)	<0.0001
Dynamic Endurance Extension	SET* vs SMT	<0.0001	<0.00001	<0.0001	1.09 (0.80, 1.39)	<0.0001
	SET* vs HEA	<0.0001	<0.00001	<0.0001	1.19 (0.88, 1.50)	<0.0001
Isometric Strength Flexion	SET* vs SMT	**0.049**	0.057	0.063	0.35 (-0.00, 0.69)	NS
	SET* vs HEA	0.993	1.0	0.994	-0.02 (-0.38, 0.34)	NS
Isometric Strength Extension	SET* vs SMT	0.057	0.067	0.073	0.34 (-0.01, 0.68)	NS
	SET* vs HEA	**0.07**	**0.087**	**0.092**	0.34 (-0.02, 0.70)	**<0.05**

* Where HEA represents home exercise and advice, SMT represents spinal manual therapy, SET represents specific exercise therapy, NS represents non-significant values (at $p<0.05$). Results in bold differ from the published results. The intervention as given by the asterisk, denotes the active comparator in the comparison, e.g. mean change SET minus mean change SMT; CI represents Confidence Interval.

Almost all the differences between arms in individual outcomes were statistically significant in favour of exercise versus comparator arms. However, for standardised isometric flexion, there were no statistically significant differences between any arms, and for standardised isometric extension there was not a statistically significant difference between exercise and manual therapy arms.

4.3.2.2.1.3 Replicating the Primary Outcome Analysis

The primary outcome analysis was performed using an ANCOVA with Tukey's post hoc analysis and baseline values as covariates. This was performed on the matched secondary outcomes (with Bonferroni adjustment). All matched outcomes demonstrated a statistically significant between-arm difference in favour of exercise, and almost all of the matched outcomes demonstrated greater SMDs than the primary outcome of pain. Only two outcomes comparing specific exercise to spinal manual therapy did not produce an SMD greater than pain (isometric strength flexion and extension) but both of these values were still statistically significant in favour of the specific exercise arm. The results are summarised in Table 4-6.

Table 4-6: ANCOVA analysis of matched outcomes in Bronfort et al. (2011)

Adjusted standardised mean difference (12-week follow-up)		
Arm comparison	**Standardised Mean Difference (95% CI)**	**Significance**
Primary Outcome Standardised Pain		
SET (vs SMT)	0.21 (-0.07, 0.5)	NS
SET (vs HEA)	0.43 (-1.03, 2.23)	NS
Standardised Static Endurance Flexion		
SET vs SMT	0.57 (0.31, 0.83)	<0.0001
SET vs HEA	0.44 (0.17 ,0.72)	<0.0001
Standardised Static Endurance Extension		
SET vs SMT	0.32 (0.08, 0.57)	0.004
SET vs HEA	0.40 (0.14, 0.65)	0.001
Standardised Dynamic Endurance Flexion		
SET vs SMT	0.59 (0.34, 0.83)	<0.0001
SET vs HEA	0.65 (0.40, 0.91)	<0.0001
Standardised Dynamic Endurance Extension		
SET vs SMT	0.84 (0.61, 1.07)	<0.0001
SET vs HEA	0.91 (0.67, 1.16)	<0.0001
Standardised Isometric Strength Flexion		
SET vs SMT	0.20 (0.01, 0.38)	0.03
SET vs HEA	-0.01 (-0.20, 0.18)	1.00
Standardised Isometric Strength Extension		
SET vs SMT	0.19 (0.00, 0.37)	0.04
SET vs HEA	0.17 (-0.2, 0.36)	0.10

Shaded cells represent primary outcome; SET is specific exercise therapy, HEA home exercise and advice, SMT spinal manual therapy. CI is confidence interval. P is significant at 0.05, and Bonferroni post hoc tests were used.

4.3.2.2.1.4 Linear Mixed Models

A linear mixed model (replicating the method of analysis for the primary outcome at 12-weeks follow-up) was also applied to each matched secondary outcome. The results are summarised in Table 4-7 in comparison to the standardised primary outcome results. All values except those for standardised isometric flexion strength produced a between-arm effect estimate that was statistically significant in favour of the specific exercise arm. Further, the SMDs for standardised dynamic endurance flexion and extension were greater than the SMD for the standardised primary outcome in both the specific exercise arm compared to the spinal manual therapy arm and specific exercise compared to

the home exercise arm. In standardised static endurance extension and flexion, the SMD was greater in the specific exercise arm than the spinal manual therapy arm but not the home exercise arm.

Table 4-7: Linear mixed model results of standardised outcomes in Bronfort et al. (2011)

	Adjusted mean difference (95% CI)		
12-week follow-up	**Effect estimate (95% CI)**	**t score**	**Sig.**
Standardised Pain (primary outcome - trial results)			
SET (vs SMT)	0.21 (-0.07, 0.5)		Not significant
SET (vs HEA)	0.43 (-1.03, 2.23)		Not significant
Standardised Static endurance Flexion			
SET vs SMT	0.55 (0.32, 0.79)	4.606	<0.0001
SET vs HEA	0.40 (0.16, 0.65)	3.228	0.001
Standardised Static endurance Extension			
SET vs SMT	0.31 (0.09, 0.52)	2.815	0.005
SET vs HEA	0.36 (0.14, 0.58)	3.174	0.002
Standardised Dynamic Endurance Flexion			
SET vs SMT	0.56 (0.34, 0.78)	4.965	<0.00001
SET vs HEA	0.63 (0.40, 0.86)	5.41	<0.00001
Standardised Dynamic Endurance Extension			
SET vs SMT	0.84 (0.62, 1.05)	7.635	<0.00001
SET vs HEA	0.92 (0.70, 1.14)	8.098	<0.00001
Standardised Isometric Strength Flexion			
SET vs SMT	0.15 (-0.00, 0.31)	1.932	0.54
SET vs HEA	-0.003 (-0.16, 0.16)	-0.033	0.973
Standardised Isometric Strength Extension			
SET vs SMT	0.17 (0.02, 0.32)	2.17	0.031
SET vs HEA	0.18 (0.02, 0.34)	2.21	0.028

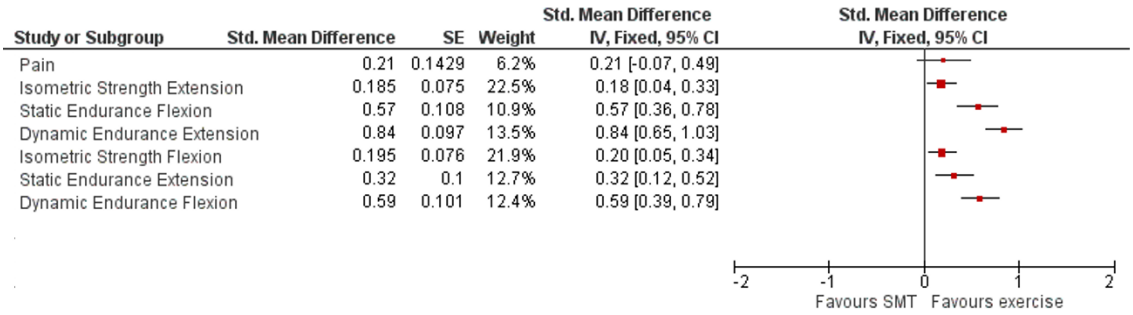
* Where HEA represents home exercise and advice, SMT represents spinal manual therapy, and SET represents specific exercise therapy; Grey shaded blocks demonstrate the original primary outcome; Positive values favour the intervention group.

4.3.2.2.1.5 Summary

The results of the secondary analysis of this RCT (117) demonstrated that the matched standardised secondary outcomes were statistically significant in favour of the specific exercise group in comparison to both control arms, and had SMDs greater than the primary outcome (difference in pain at 12-weeks) in all results apart from the standardised isometric flexion and extension strength in the ANCOVA analysis presented in Figure 4-1. However, although the SMD effect

estimates were greater in the secondary outcomes, there was only a statistically significant difference between dynamic endurance extension and the primary outcome of pain.

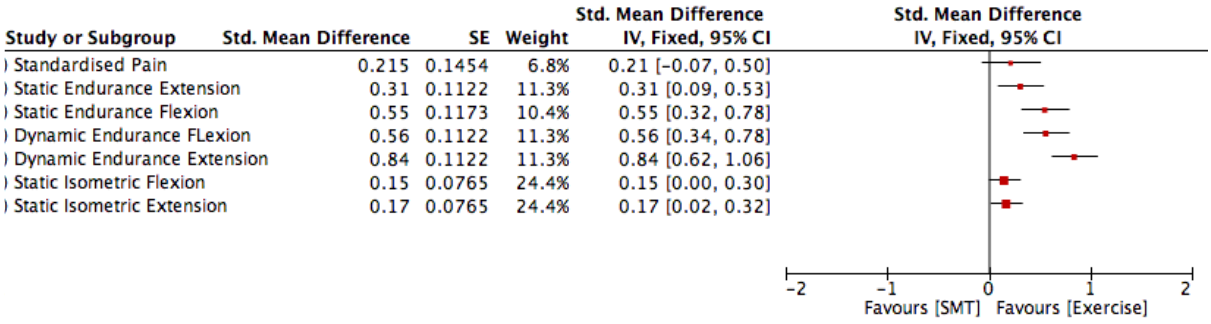
Figure 4-1: Forest plot to demonstrate size and direction of effect of outcomes in Bronfort et al. (2011) using ANCOVA



Where IV inverse variance, CI is confidence intervals, SE is standard error, SMT is spinal manual therapy, Std. is standard as part of SMD.

Similarly, five of seven results showed statistically significant between-arm differences, with the greater SMDs favouring specific exercise in comparison to spinal manual therapy. The two exceptions were the outcomes of standardised isometric flexion (SMD and non-significant results) and extension strength (SMD only) in the linear mixed model results in Figure 4-2. As with the ANOVA results, there was only a statistically significant difference between dynamic endurance extension and the primary outcome of pain SMDs.

Figure 4-2: Forest plot to demonstrate size and direction of effect of outcomes in Bronfort et al. (2011) using linear mixed models



IV inverse variance, CI is confidence intervals, SE is standard error, SMT is spinal manual therapy, Std. is standard as part of SMD.

When comparing the size of the SMDs across outcomes, all were greater for the matched secondary outcomes than the reported primary outcome (pain) at 12-weeks in the ANCOVA and linear mixed model analysis apart from isometric flexion and extension strength. In the ANCOVA analysis, all of the matched outcomes demonstrated statistically significant results in favour of specific exercise over spinal manual therapy. Using the linear mixed model of analysis, five of the six matched outcomes were statistically significant in favour of specific exercise, apart from standardised isometric flexion strength. In both analyses, dynamic endurance extension was the only secondary outcome to demonstrate a statistically significant difference with the primary outcome in favour of supervised exercise.

4.3.2.2.2 *Secondary Analysis of Groessl et al.* (127)

The Groessl et al. RCT (127) included four different time-points (baseline, 6-weeks, 12-weeks, and 24-weeks) and four secondary outcomes matched their

stated treatment targets (strength, flexibility (two measures) and pain relief). The primary outcome was physical function, and there was no statistically significant difference in physical function between yoga (exercise intervention) and waiting list (comparator) at the first two follow-ups (primary time-point was 12-weeks), but there was a statistically significant difference at 24-weeks in favour of the yoga group.

4.3.2.2.2.1 Data Modification

On reviewing the matched secondary outcomes of interest (four outcomes) via descriptive variables, there were no missing data identified. The outcome "extension" scored on a negative scale, requiring multiplication by minus one to score in the same direction as the other scales. The dataset was converted from a wide to a long format as described in section 4.2.3.2.1 by merging all four time-points included for each targeted outcome.

4.3.2.2.2.2 Replicating the Original Results: Linear Mixed Model Analysis of Pain

It was not possible to repeat the analysis performed on the primary outcome as this was not requested in the dataset (as it was not one of the matched targets). The replicated linear mixed model analysis of the published secondary outcome measure of pain gave similar results to those of the original (published) results (see Table 4-8). Pain was the only targeted secondary outcome that was reported in the published paper via the linear mixed modelling approach, but, other targeted secondary outcome measures were also available in the requested dataset for analysis (results provided later within Table 4-9).

Table 4-8: Linear mixed model analysis of the original trial results in comparison to the replicated secondary outcome (pain) results in Groessl et al. (2017)

	Adjusted mean difference (95% CI)		
Yoga vs WL	**Effect estimate (95% CI)**	**t score**	**Sig.**
Pain (trial results)			
12-week follow-up	0.65 (0.2, 1.1)		0.005
This Analysis of Pain (Using covariate type UN)			
12-week follow-up	0.62 (0.2, 1.1)	2.74	0.007

*Where WL is waiting list, CI is confidence intervals, Sig is significance at 0.05. All outcomes represent mean yoga minus mean waiting list such that negative values favour yoga over the control arm.

4.3.2.2.2.3 Linear Mixed Model Analysis

The individual secondary outcomes tabulated in Table 4-9 demonstrated that only pain showed statistically significant differences in favour of the exercise arm at the primary time-point. Although standardised pain, plank and flexion ROM SMDs were greater than those of the primary outcome estimates at 12-weeks, standardised extension ROM was smaller than the primary outcome SMD at 12-weeks.

Table 4-9: Linear mixed model analysis of the secondary outcome results in Groessl et al. (2017)

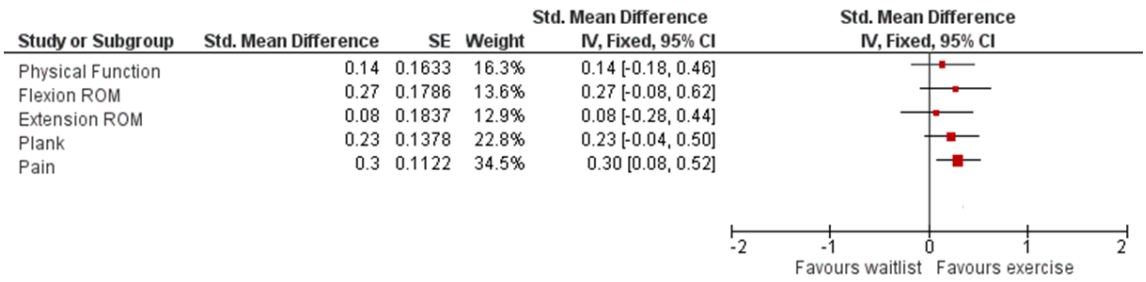
	Adjusted mean difference (95% CI)		
Yoga vs WL	**Effect estimate (95% CI)**	**t score**	**Sig.**
Standardised Primary Outcome: Physical Function (trial results)			
12-week follow-up	0.14 (-0.18, 0.46)		0.340
Standardised Pain			
12-week follow-up	0.30 (0.08, 0.52)	2.741	0.007
Standardised Plank			
12-week follow-up	0.23 (-0.04, 0.51)	1.64	0.105
Standardised Flexion ROM			
12-week follow-up	0.27 (-0.08, 0.61)	1.538	0.127
Standardised Extension ROM			
12-week follow-up	0.08 (-0.28, 0.44)	0.456	0.649

Grey shaded blocks demonstrate the original primary outcome; ROM represents range of movement. Where WL is waiting list, CI is confidence intervals, Sig is significance at 0.05. All outcomes represent mean yoga minus mean waiting list such that a positive value favours yoga over the waiting list.

4.3.2.2.2.4 Summary

The results of this analysis demonstrated that only one (pain) of the four matched secondary outcomes had a greater SMD and statistical significance than the original unmatched primary outcome (physical function) at the primary time-point (i.e. at 12-weeks follow-up). However, three of the four matched secondary outcomes [standardised pain (SMD 0.30 (95% CI 0.08, 0.52)), standardised plank (SMD 0.23 (95% CI -0.04, 0.51)), flexion range of motion (SMD 0.27 (95% CI -0.08, 0.61))] demonstrated greater SMDs (without statistical significance) than the primary outcome (physical function) (SMD 0.14 (95% CI -0.18, 0.46)) at the primary time-point, as seen in Figure 4-3, although these were not statistically significantly different to the primary unmatched outcome.

Figure 4-3: Forest plot to demonstrate size and direction of effect of outcomes in Groessl et al. (2017)



Where ROM is range of motion, IV inverse variance, CI is confidence intervals, SE is standard error.

4.3.2.2.3 *Secondary Analysis of Harris et al.* (130)

The Harris et al. RCT (130) had two time-points (baseline and 52-weeks) and one matched secondary outcome (fear-avoidance beliefs). The RCT compared three groups: 1) a brief intervention (BI), 2) brief intervention with the addition of physical exercise (BI+PE) or 3) brief intervention with the addition of a cognitive

behavioural therapy component (BI+CBT). The primary outcome was return to work at 52-weeks, and the main trial results showed no statistically significant between-arm differences.

4.3.2.2.3.1 Data Modification

The matched secondary outcome of fear-avoidance beliefs (comprising two subscales of beliefs about physical activity, and beliefs about work) and physical function were checked with descriptive statistics. Values for these outcomes were within the expected ranges, and there were no missing data. All three variables trended in the same direction and thus did not require modification. The dataset was suitable to be used in wide format for further analysis.

4.3.2.2.3.2 Calculation of SMD for the Primary Outcome

The primary outcome used was the proportion of participants who returned to work (RTW), and the SMD was calculated according to the methods described in chapter 3, section 3.2.6.1. The SMD for RTW was 0.16 (95% CI -0.32, 0.00) in favour of the control arm compared to the exercise arm.

4.3.2.2.3.3 Replicating the Original Method: Chi-Square Analysis and ANOVA Analysis

The original primary outcome was analysed with chi-square analysis. This was replicated and shown below, in Table 4-10, but one participant's details appear to have been lost in the dataset provided.

Table 4-10: Results of chi-square analysis performed in Harris et al. (2017)

Reported Results				Intervention Groups		
			BI	BI+ PE	BI+CBT	Total
Primary Outcome	Returned to work	Count	60	31	30	121
		% within groups	60	51.7	54.6	-
	Total	Count	100	60	55	215
		Chi-square	1.15			
		p-value	0.563			
Analysed Results				Intervention Groups		
			BI	BI + PE	BI+CBT	Total
Primary Outcome	Not Returned to work	Count	40	29	25	94
		% within groups	40.4	48.3	45.5%	43.9%
	Returned to work	Count	59	31	30	120
		% within groups	59.6	51.7%	54.5%	56.1%
Total		Count	99	60	55	214
		% within groups	100%	100%	100%	100%
		Chi-square	1.024			
		p-value	0.599			

Where BI represents brief intervention, BI+PE represents brief intervention and physical exercise and BI+CBT represents brief intervention with cognitive behavioural therapy.

For all the secondary outcomes, the authors performed a mixed between- and within-subject analysis of variance with one between-group factor (BI, BI + PE, BI + CBT) and with one within-subjects/repeated measures factor (baseline and 12-months follow-up) (see Table 4-11). The effect of time and the interaction effect (Time X Group) were reported, and the F and p-values were compared to the reported published values, as displayed in Table 4-11.

Table 4-11: Replication results of ANOVA analysis in Harris et al. (2017)

	Time*Group Interaction		
BI vs PE vs CBT	***F*-value**	**Sig.**	**Partial Eta Squared**
Fear-Avoidance Beliefs Scores (physical activity) (reported)			
12-month follow-up	**2.270**	**0.107**	0.031
Fear-Avoidance Beliefs Scores (calculated)			
12-month follow-up	2.270	0.107	0.031

Where BI represents brief intervention, PE represents brief intervention and physical exercise and CBT represents brief intervention with cognitive behavioural therapy.

4.3.2.2.3.4 ANOVA Analysis of Standardised Results

For this secondary analysis, the standardised (treatment-targeted) secondary outcome(s) were analysed using ANOVA: the standardised physical function scores (unmatched), as well as the matched standardised fear-avoidance beliefs scores for physical activity and work (shown in Table 4-12). The only outcome to demonstrate a statistically significant difference between arms (group*time interaction using the ANOVA) was fear-avoidance beliefs relating to work (F-value$_{2,\ 142}$ 2.337, p=0.049). However, post hoc tests comparing arms did not demonstrate any statistically significant differences. Although, when comparing the mean difference between arms, a greater difference was identified in fear-avoidance beliefs about work scores when comparing the mean exercise (BI + PE) minus mean control (BI) groups (0.289) in comparison to the unmatched physical function scores (0.177).

Table 4-12: Results of ANOVA analysis performed in Harris et al. (2017)

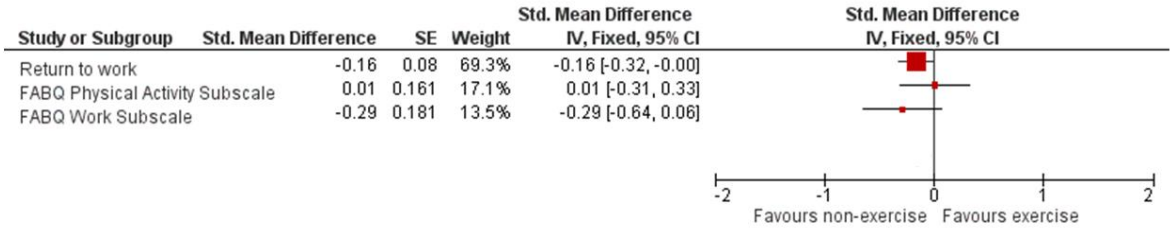
Standardised RTW Score	**Calculated SMD**		
BI + PE vs BI	-0.16 (-0.32, 0.00)		
Standardised Physical Function Scores	**Time*Group Interaction**		
BI vs PE vs CBT	***F*-value**	**Sig.**	**Partial Eta Squared**
12-month follow-up	1.118	0.330	0.016
Post Hoc Analysis (Tukey HSD)	**Mean Difference**	**Sig.**	**Std. Error**
BI + PE vs BI	0.18 (-0.2, 0.6)	0.561	0.172
BI + PE vs BI + CBT	0.28 (-0.2, 0.7)	0.315	0.193
Standardised Fear-Avoidance (Physical Activity)	***F*-value**	**Sig.**	**Partial Eta Squared**
12-month follow-up	0.952	0.388	0.013
Post Hoc Analysis (Tukey HSD)	**Mean Difference**	**Sig.**	**Std. Error**
BI + PE vs BI	0.01 (-0.4,0.4)	1.000	0.161
BI + PE vs BI + CBT	0.21 (-0.2, 0.6)	0.458	0.179
Standardised Fear-Avoidance (Work)	***F*-value**	**Sig.**	**Partial Eta Squared**
12-month follow-up	1.895	0.154	0.026
Post Hoc Analysis (Tukey HSD)	**Mean Difference**	**Sig.**	**Std. Error**
BI + PE vs BI	-0.29 (-0.7, 0.1)	0.250	0.181
BI + PE vs BI + CBT	0.03 (-0.4, 0.5)	0.992	0.199

Where RTW represents return to work, BI represents brief intervention, PE physical exercise and CBT represents cognitive behavioural therapy. All Post hoc analyses demonstrate the difference between intervention minus control arm: e.g. mean (BI + PE) minus mean (BI) (in BI + PE vs BI) or BI + PE minus BI + CBT (in BI +PE vs BI + CBT). Positive mean difference values indicate favour of the intervention arm.

4.3.2.2.3.5 Summary

This analysis demonstrates that the use of secondary matched outcomes did not produce statistically significant results in favour of the exercise arm. When assessing SMD results, the matched secondary outcome of fear-avoidance beliefs about physical activity generated a greater SMD in favour of the exercise arm (BI + PE) in comparison to the brief intervention alone (SMD 0.01 (95% CI -0.31, 0.33) than the unmatched primary outcome (SMD -0.16 (95% CI -0.32, 0.00), as seen in Figure 4-4, but these differences were not statistically significant.

Figure 4-4: Forest plot to demonstrate size and direction of effect of each outcome in Harris et al. (2017)



Where FABQ is the fear-avoidance beliefs questionnaire, IV inverse variance, CI is confidence intervals, SE is standard error, Std. is standard as part of SMD.

4.3.2.3 **Summary of Results**

Out of the five RCT datasets included in this secondary analysis study, three had greater SMDs and statistical significance in favour of exercise compared to control interventions when a matched secondary outcome was used in comparison to an unmatched primary outcome (see summary Table 4-13) (117,127,140). Tilbrook et al. (142) was the only RCT of the five re-analysed to produce greater between-arm differences in favour of exercise when using an unmatched primary outcome. Of the three RCT datasets analysed, two demonstrated larger, statistically significant effects in favour of the exercise arm with at least one matched secondary outcome at the primary time-point(s), compared to an unmatched primary outcome (117,127). The Harris et al. (130) dataset did not demonstrate any statistically significant between-arm differences using any of the outcomes, but the use of the matched secondary outcome generated a greater SMD in favour of the exercise group than when using the unmatched primary outcome.

Table 4-13: Summary of SMD results for each included RCT and outcome

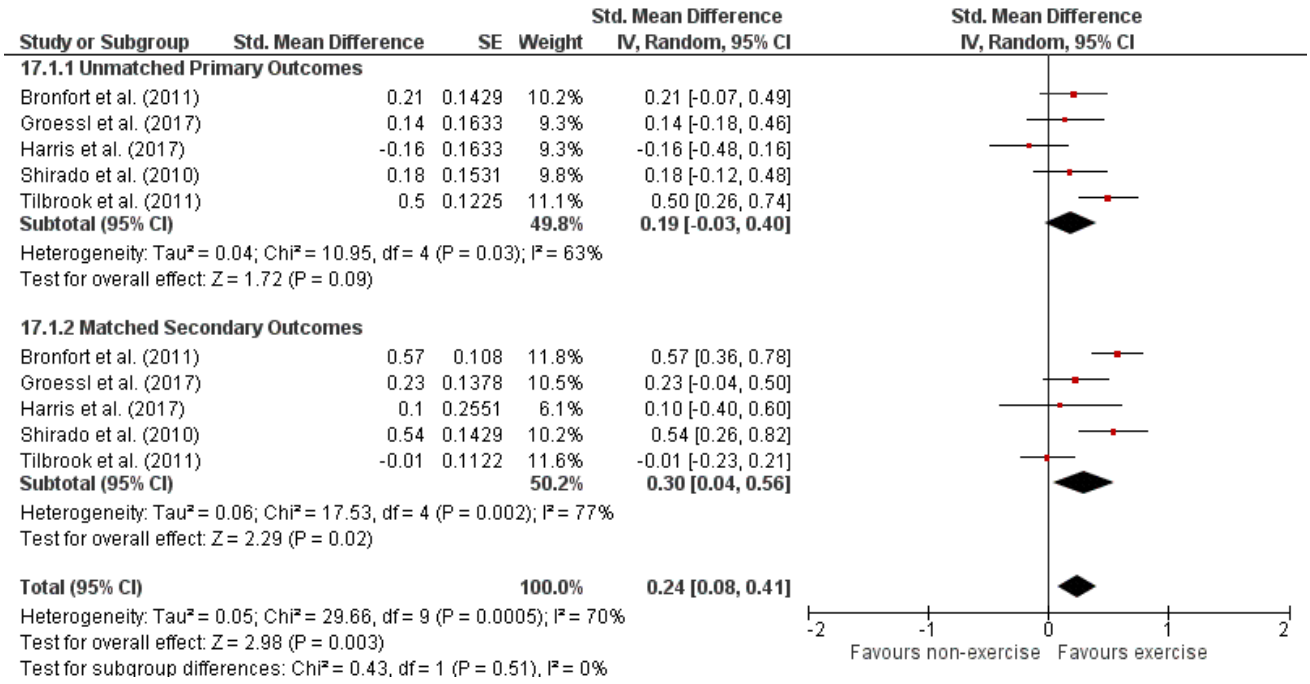
Trial	**Comparator**	**Outcome (Primary outcome shaded in grey)**	**SMD (95% CI)** Primary time-point	**Conclusion Changed?**
Bronfort et al. (117)	Ex vs SMT	Standardised Pain	0.21 (-0.07,0.50)	Yes
		Standardised Static Endurance Flexion	0.57 (0.31, 0.83)	
		Standardised Static Endurance Extension	0.32 (0.08, 0.57)	
		Standardised Dynamic Endurance Flexion	0.59 (0.34, 0.83)	
		Standardised Dynamic Endurance Extension	0.84 (0.61, 1.07)	
		Standardised Isometric Strength Flexion	0.20 (0.01, 0.38)	
		Standardised Isometric Strength Extension	0.19 (0.00, 0.37)	
Groessl et al. (127)	Ex vs WL	Standardised Physical Function	0.14 (-0.46, 0.18)	Yes
		Standardised Pain	0.30 (0.08,0.52)	
		Standardised Plank	0.23 (-0.04, 0.51)	
		Standardised Flexion ROM	0.27 (-0.08, 0.61)	
		Standardised Extension ROM	0.08 (-0.28, 0.44)	
Harris et al. (130)	BI +PE vs BI	Standardised Return to Work	-0.16 (-0.32, 0.00)	No
		Standardised Fear-Avoidance Beliefs (Work)	-0.29 (-0.7, 0.1)	
		Standardised Fear-Avoidance Beliefs (Physical Activity)	0.01 (-0.4,0.4)	
Shirado et al. (140)	Ex vs NSAIDS	Standardised Pain	0.17 (-0.12,0.47)	Yes
		Standardised Physical Function	0.27 (-0.02, 0.55)	
		Standardised HRQoL	0.29 (-0.00, 0.57)	
		Standardised Forward Finger Distance	0.54 (0.26, 0.83)	
Tilbrook et al. (142)	Ex vs UC	Standardised Physical Function	0.50 (0.26; 0.74)	**Yes**
		Standardised Pain	-0.01 (-0.23, 0.22)	

Positive values favour the exercise intervention; negative values favour the control intervention. Grey shaded boxes represent the standardised primary outcome. ROM is range of motion, HRQoL is health-related quality of life; Ex is exercise, SMT is spinal manual therapy, WL is waiting list, PE is physical exercise, BI is brief intervention, NSAIDS is non-steroidal anti-inflammatories; UC is usual care; FU is follow-up. Bold values represent changes not in favour of exercise group.

Figure 4-5 showed an overall effect between the unmatched primary outcome and the first-mentioned matched outcomes of the analysed trials. As can be seen from the unmatched primary outcomes, the pooled effect (SMD 0.19 (95% CI -

0.03, 0.40) p=0.09) is smaller than that of the first-mentioned matched secondary outcome (SMD 0.30 (95% CI 0.04,0.56) p=0.02). As the same trials are included, one would expect the heterogeneity values to be similar in both meta-analyses; however, the first analysis represents moderate heterogeneity (63%), whereas the matched secondary outcomes analysis represents high heterogeneity (77%) (106). These forest plots only represent the first-mentioned matched outcomes at the primary time-point (or the time-point closest to the end of the exercise intervention) as it is recommended that only one-time point for the same sample is incorporated in a meta-analysis (168). There was no statistically significant difference between the sub-groups (primary outcome compared to the first matched outcome) based on the p-value (p=0.51). The between-group SMD when calculated was 0.11 ((95% CI -0.34, 0.57) (p=0.51)). Only three of the graphed SMD outcomes changed: two from non-significant to statistically significant (117,140) and one from statistically significant to non-statistically significant (142).

Figure 4-5: Forest plot to demonstrate the pooled effect of the SMD for unmatched primary outcomes in comparison to matched secondary outcomes



Std. is standard as part of SMD, SE is the standard error, IV is inverse variance, CI is confidence interval.

4.4 Discussion

This exploratory secondary analysis demonstrated in three of the five RCTs analysed (117,127,140), and a fourth showed potentially (130) that using an outcome matched to the treatment targets of exercise produced larger SMDs, when exercise is compared to a non-exercise control, than when using an unmatched primary outcome, although this difference was not statistically significant between groups. In one RCT, greater between-arm SMDs were observed in favour of the control arm with a matched secondary outcome in contrast to the primary outcome (130), but these were not statistically significant. Only one RCT found greater between-arm differences with an unmatched

primary outcome in favour of the exercise arm (142). The results of this exploratory analysis of five RCTs suggests that the use of an outcome that better matches the treatment targets of exercise can result in greater SMDs, more statistically significant between-arm differences, and different conclusions reached by RCTs in favour of exercise in comparison to a control arm, than when an unmatched primary outcome is used. It is, therefore, possible that the results and conclusions of exercise RCTs may alter if outcomes are better matched to exercise treatment targets.

This results of this secondary data analysis adds support to those from the earlier systematic review that matching outcomes to the treatment targets of the intervention may result in greater effect sizes than when using an unmatched outcome. It is important to note that exercise interventions were favourable across nearly all outcomes to varying degrees with the size of the effect varying from very small (<0.1) to large (>0.8) but mostly lying within the 'small-to-medium' range of 0.2 to 0.3 (107)), in line with what has previously been observed in reviews of exercise for LBP (26,27,29). Only one RCT reported results that favoured the non-exercise control group (130). The overall SMD of the matched outcomes from these five RCTs was the same size ((0.30 (95%CI 0.04, 0.56) p=0.02) in favour of exercise) as that reported in recent literature of exercise interventions in comparison to a brief intervention or minimal control ((0.32, 95% CI 0.19, 0.44, p<0.01) in favour of exercise)(27). However, the unmatched primary outcomes grouped SMD was even smaller than this result (0.19 (95% CI -0.03, 0.40) p=0.09), despite being mostly compared to a minimal intervention (only Bronfort et al. (117) included an active comparator arm in these trials). This

provides exploratory evidence that matching the primary outcome of an RCT to the treatment targets of the exercise intervention may influence the effect size of the between-arm difference and, therefore, may alter the conclusions of the RCT.

4.4.1 Outcome Domains used by RCTs Included in this Analysis

One of the challenges of performing this secondary analysis was the wide variety of primary outcome domains used across these five RCTs, with the primary outcome domain selection tending to reflect recommendations in the core outcome domain LBP literature (78,82): three RCTs used physical function measures, two used pain intensity measures, one reported return to work rates and one reported health-related quality of life[10]. The CONSORT statement has a category for the explanation of the proposed rationale for the trial (category 2a); however, there is no mention of whether this rationale matches the treatment targets of the interventions or the outcomes utilised (category 6a)(63). There are many reasons for the selection of primary outcome domains in RCTs, such as the recommendations regarding core outcome sets for RCTs, funders' requirements, patient preferences, ease of measurement, ability to compare results with other RCTs, etc. However, the results of this analysis suggests that the selection of the primary outcome should also be guided by the rationale for the specific intervention (71). Further, the selection of the primary outcome specifically impacts on the required sample size, the desired MCID and the

[10] These numbers do not total 5 as one trial specified 3 primary outcome domains (Shirado et al. (140)).

conclusions of the RCT. Thus, ensuring the most appropriate domain and measure selected is paramount.

The number of matched outcomes to exercise intervention targets varied across the five RCTs, with three including only one matched outcome, while the other two included four and six matched outcomes respectively. This demonstrates the varying exercise intervention targets stated and measured by trial authors. It is clear from this study that there is no consensus on what the treatment targets of exercise interventions are for patients with persistent NSLBP. This could be due to inadequate exercise programme theory, as only recently has greater awareness and emphasis been placed on intervention logic models and programme theory to demonstrate the proposed mechanisms of the effect of interventions (52).

4.4.2 Data Analysis Method: Linear Mixed Models Compared to ANOVA

Although varying outcome measures were used in the sample of five RCTs in this study, they were all standardised within the individual datasets wherever possible. However, the analysis method varied across RCTs. Many of the more recent RCTs used linear mixed modelling approaches which incorporate all time-points in regression analysis, in contrast to the ANOVA or ANCOVA approach. This approach was not always possible to replicate in the secondary analyses due to the sometimes limited number of follow-up time-points, such as in Bronfort et al. (117), where patient-reported outcomes were captured at four time-points, but physical performance outcomes were only captured at two time-points.

However, two trials used the ANOVA approach, with or without a covariate to control for baseline values (117,130). Linear mixed models are often preferred over ANOVA due to their ability to control for missing data and unbalanced datasets (169). For the results of the ANOVA to be valid, this requires the assumptions of compound symmetry and complete data to be met (169). Compound symmetry means that the variances and the covariance's of the repeated measures are similar, and if this assumption is violated, then this increases the risk of a Type I error, but this can be accounted for by adjusting the degrees of freedom (169). This was not performed in this secondary analysis which increases the risk of Type I error, but given this was the preferred method by the RCT authors, it was utilised as per all other analyses (117,130).

In the meta-analysis of the summary values, one would expect the heterogeneity values to be similar in both meta-analyses given that the same trials are included. However, the first analysis represented moderate heterogeneity (63%), whereas the matched secondary outcomes represented high heterogeneity (79%) (106). This difference in heterogeneity values may be due to the use of SMD values across a variety of different outcome measures – the first analysis compared the primary unmatched outcomes of pain (117,140) physical function (127,142) and return to work (130), whereas the second analysis compared measures of flexibility (140), pain (127,142), strength (117), and fear-avoidance beliefs (130).

4.4.3 Process Evaluation of Exercise Interventions in the Included RCTs

Process evaluation is an important component of investigating complex interventions (15,52), helping to make explicit the theory underlying the intervention, assess how the intervention was delivered, whether the targets of the intervention were achieved, and how the context may have affected the delivery. In this study, it was difficult to assess to what degree treatment targets were delivered (and whether some aspects were not mentioned in the published reports but nevertheless were still delivered), as although these RCTs specified their exercise treatment targets, there may have been elements of underlying theory not mentioned (such as the fear-avoidance model) or other treatment targets that were not specified in the published text. Groessl et al. (126) targeted strength, flexibility (two measures) and pain, and found greater effect estimates at the primary time-point in three of these four matched targets than with their primary outcome (physical function). Tilbrook et al. (142) similarly targeted their yoga intervention at pain, mobility and strength, despite pain being the only outcome they measured and reported. Although targets may be defined by the authors, they may not be equally implemented in intervention delivery or captured in the selection of outcomes. This may be due to the level of tailoring versus standardisation of the intervention (15,52), which may affect the weighting of the delivery of the intervention targets. Without adequate process evaluation or intervention description, it is difficult to assess what impact the context had or the mechanisms of action of the intervention (if unexpected results are obtained) (52). Descriptions of complex interventions have generally improved since the

advent of the TIDieR guidance (64), but this may not always be helpful if the exercise is not described in relation to the treatment targets, or to a degree that is replicable in clinical practice (170). In a review of LBP RCTs, only 13% were judged to be reproducible in clinical practice (161).

It may be important to additionally consider not only what exercise intervention targets were stated by the RCT authors, but also the dosage and adherence levels of the participants within the intervention groups, as well as the type of control comparison. For example, both Groessl et al. (127) and Tilbrook et al. (142) used yoga as their intervention in their RCTs, yet each described slightly different intervention targets. Both found similar effects on pain outcomes, but the effects on physical function were very different (SMD of 0.15 in Groessl et al. RCT (126) vs 0.50 in Tilbrook et al. RCT (143)) raising the question as to why this may be and what differed between the delivery of these two apparently similar interventions. Although Groessl et al. (126) had compared yoga with a waitlist control, it appears they both compared the yoga arm to a usual care arm. Ironically, although Tilbrook et al. (143) reported a greater effect of yoga versus usual care on physical function at 12-weeks, the yoga intervention in the RCT by Groessl et al. (126) offered almost double the dosage of yoga than in the Tilbrook RCT. In Tilbrook et al. (143) the intervention was once weekly for 75 minutes over 12 weeks (12 classes), whereas in Groessl et al. (126) the intervention was twice weekly for one hour over the same period (24 classes). The mean number of yoga classes attended by participants in Groessl et al. (126) was 12.3 classes (58%). In the Tilbrook et al. RCT (143), adherence levels differed, requiring the attendance of three of the first six classes and any three thereafter, which 60%

of participants achieved. Thus, there does not appear to be a simple relationship between adherence or dosage and between-arm differences in physical function in these two RCTs, suggesting that other factors may have played a role in generating the between-arm differences of these similar exercise interventions. Previous work suggests that simply enrolling LBP patients in RCTs (or other studies) may provoke responses reflective of the nonspecific effects of care-seeking and treatment (7). Without clear process evaluation, it is difficult to identify the key factors that may have played an important role in generating such different between-arm differences, as seen in these two trials. Adherence, duration and dosage of the exercise do not appear to adequately explain the differences in outcomes of these two RCT examples. Therefore, more in-depth analysis such as process evaluation, including *a priori* specified mediation analysis and qualitative evaluation of the fidelity of the interventions delivered may provide valuable guidance for future research.

4.4.4 Strengths and Limitations of This Analysis

This is the first attempt to explore whether matching the outcome in RCTs of exercise for persistent NSLPB to the treatment targets of the exercise interventions might change the results and conclusions of RCTs. This work utilised the RCTs identified in the previous systematic review (chapter 3) and complements the work on core outcome sets for LBP RCTs (71,78,82). The use of existing RCT datasets and the close replication of the original RCT data analyses are further strengths of this work. In addition, the way the SMDs were derived, standardising the outcome measure scales to allow direct comparison

of the size of the effects across the different outcome measures used in the previous RCTs is a strength.

Limitations include the small number of RCT datasets included (determined by the RCTs identified by the earlier systematic review that met the criteria for inclusion in this secondary analysis study), but this work provides initial exploratory evidence that support the need to consider better matching of the outcomes in RCTs to the intervention targets. Another limitation is the sheer number of additional secondary analyses per trial, which increases the likelihood that a statistically significant finding is found when one isn't actually present (Type I error)(171).

Two of the available RCT datasets did not include all the unmatched primary outcomes described in the paper and, thus, these analyses were not able to be replicated which was an oversight at the time of requesting data. In order to request data, a fully worked plan for how the data was to be used was developed, in order to clearly detail what was required and why. However, in future, the inclusion of the unmatched variables would have made further analysis easier and more transparent. In future, requesting a complete dataset with all potentially relevant variables and time-points may be advantageous to prevent delays and the limitations identified in this analysis happening in the future. This study was an exploratory analysis to assess the impact of using a matched secondary outcome in place of an unmatched primary outcome. Secondary analyses of existing data are less frequently used in research but, are more time-efficient, resource-economic and cost-effective (172). It is, however, important to note that

none of these trial datasets were initially collected with the intention to assess whether the matched secondary outcomes would generate a greater SMD and statistical significance than the original primary outcome (172). Further, obtaining suitable datasets can be a challenge, as noted in the process of acquiring these datasets (three of the four requested datasets were obtained). These datasets were identified from the previously performed systematic review (chapter 3). The search strategy had no lower time limit, allowing for inclusion of trials that may have been performed more than twenty years prior, and thus limiting the ability to access the datasets or contact the authors.

4.4.5 Implications for Further Research

The results of this study provide exploratory evidence that matching outcomes to the targets of exercise interventions for persistent NSLBP may alter the results and conclusions of exercise RCTs, although, further research is needed. This was an exploratory study analysing five previous RCTs, and further research should be conducted that examines this issue for exercise and persistent NSLBP, exercise for other conditions, and indeed more broadly for other complex interventions before stronger conclusions can be made. In addition, process evaluation should be considered prior to intervention development and specification in future trials, to ensure adequate consideration of mechanisms of action and treatment targets. This should occur throughout the trial delivery as well as on conclusion of the RCT, to evaluate whether the proposed intervention was delivered as expected and whether any other theory should be added to the intervention logic model. For future exercise RCT designers, the implications are

that: they should make clear what their exercise treatment targets are, report these in the trial protocols and papers, and include appropriate outcome measures that can capture the change in these targets over time.

This study further underlines the need to agree on the treatment targets for exercise interventions in persistent NSLBP. This would be helpful to facilitate future mediation analysis in RCTs, formally testing the variables through which the exercise intervention is believed to have effects. When there are multiple possible matched outcomes that could be used in an RCT as the nominated primary outcome, and without clear intervention programme theory to understand how the intervention might affect these identified treatment targets, it may distort the overall purpose of the RCT to select only one of these outcomes (85). Watt et al. (84) suggest that composite outcomes, comprised of the most appropriate individual outcomes, may improve the statistical precision of the trial. As seen in this chapter, many RCT authors' stated more than one treatment target of the exercise intervention, and these were captured by different outcomes. Thus, a composite outcome composed of these multiple matched outcomes may be more responsive than a single matched outcome. This is explored further in chapter 5.

4.5 Summary

The results of this chapter, which focuses on secondary analysis of five RCTs, provides exploratory evidence that better matching of the outcomes in RCTs to the treatment targets of exercise for NSLBP, may alter the results and

conclusions reached by these trials. When outcomes are used that are matched to the treatment targets of exercise, RCTs may show greater between-arm effect sizes and more statistically significant results, in favour of exercise. The following chapter seeks to explore the impact of composite outcomes on the effect sizes and statistical significance reported in exercise RCTs.

5 Chapter 5: Exploratory Development of Composite Outcomes in Exercise RCTs for NSLBP

Summary

This chapter reports the secondary analyses of four RCT datasets, wherein a composite outcome, comprised of the matched outcomes, was created and analysed using the same method as the primary outcome. The results of the composite outcomes were compared to the primary outcomes, to explore whether the results and conclusions of the four RCTs would have altered had a matched composite outcome been used.

Excerpts of this chapter have been written up as part of a paper:

Wood L, Foster NE, Lewis M, Bronfort G, Groessl E, Hewitt C, Miyamoto G, Reme SE, Bishop A. Matching the outcomes to treatment targets of exercise for low back pain: does it make a difference? Results of secondary analyses from individual patient data of randomised controlled trials and pooling of results across trials in comparative meta-analyses. Under review.

5.1 Introduction

Chapter 4 described an exploratory secondary analysis study of five previous RCTs testing exercise for NSLBP. The results showed that better matching the outcomes to the treatment targets of the exercise intervention can alter the results and conclusions of NSLBP RCTs, with most (3/5) RCTs concluding a

greater between-arm SMD in favour of the exercise intervention. However, Watt et al. (85) suggest that nominating one single primary outcome, in an RCT of a complex intervention where the intervention has a range of different potential outcomes, may distort the overall purpose of the RCT. Composite outcomes, including two or more component outcomes (173), maybe more suitable than single primary outcomes in RCTs of complex interventions, such as exercise, and maybe better able to demonstrate the effects of the intervention on a range of key treatment targets. The aim of this study was, therefore, to understand whether using a composite outcome in RCTs testing exercise for persistent NSLBP might alter the results and conclusions of these RCTs, when the composite outcome is composed of multiple individual outcomes matched to the authors' stated exercise treatment targets.

5.1.1 Aims and Objectives

Aim: This secondary analysis aimed to explore whether the creation of a composite primary outcome matched to the treatment targets of exercise in persistent NSLBP RCTs alters the results and conclusions of the RCTs.

Objectives:

i. To replicate the analysis applied to the primary outcome(s) by the authors of the RCTs or, where these were not available, on the secondary outcome(s).

ii. To calculate a composite outcome standardised mean difference (SMD) using standardised averages of outcomes matched to the treatment targets of the exercise intervention per identified RCT.

iii. To calculate a co-primary composite outcome SMD using standardised averages of the primary outcomes where nominated by the trials' authors.

iv. To compare composite and co-primary composite outcome SMDs with the reported primary outcome SMDs per identified RCT.

5.2 Methods

5.2.1 Sample

Identified RCTs from the "partially matched" and "matched" category classified in the systematic review (chapter 3, section 3.3.6) were included where more than one outcome domain matched more than one specified exercise treatment target. There was no pre-specified target number of datasets as this was exploratory analyses based on the results of the systematic review categorisation.

5.2.2 Data Extraction

Pertinent information to inform this data analysis was already extracted as part of the systematic review (chapter 3) as follows:

i. The treatment targets of the exercise interventions

ii. The primary and secondary outcome(s) measured and reported within the RCTs

iii. The outcomes that matched the authors' stated exercise treatment targets

iv. The type of analysis performed using the outcomes

5.2.3 Data Analysis

A composite outcome composed of the averaged standardised matched outcomes within each RCT was developed using individual participant data. This ensures the precision of the composite measure. For each included RCT the information used was dependent on the identified treatment targets and the matched outcome domains extracted in the systematic review earlier in this thesis (chapter 3). Data extraction of the treatment targets and outcome domains was performed by pairs of two independent reviewers as described in chapter 3, section 3.2.4.

Some of the methods for combining outcomes into a composite outcome aim to improve domain coverage, whilst others aim to improve responsiveness (174). Optimally, the method used should do both (166). A variety of methods to produce a relevant composite outcome exist, simple averaging of standardised scores, weighted averaging, principal component analysis and meaningful grouping; of which simple averaging is the most commonly used (175). The simplest means to attaining a composite outcome of continuous variables is through averaging the included outcomes; however, this assumes that equal importance is placed on each outcome, and that units from each outcome are comparable (166,175). This method intends to preserve the distribution of the raw scores. Although weighted averaging is another method reported for calculating composite outcomes, in the analyses reported in this chapter this would have been difficult to perform, as the percentage weight allocated to each

targeted outcome would depend on the degree it was targeted within the exercise intervention, and information about this was not available in the trial reports.

Given the above, exploratory analyses were performed on the available RCT datasets by creating a composite variable at baseline and the primary follow-up time-points using simple averaging of standardised variables (175). This varied per RCT, where each composite was composed of the outcomes that matched the treatment targets that the authors' described. The type of analysis mirrored that used in the published reports of each RCT.

In one dataset the primary analysis method was analysis of variance with a covariate (ANCOVA), and thus the method above was used to create a composite which was also analysed with ANCOVA, as well as the Multivariate ANCOVA (MANCOVA) method. In the other three included RCT datasets, the linear mixed model method was used in their primary analyses, and this was then used to create a composite in these three datasets. The linear mixed model includes data from all time-points captured, whereas the ANCOVA was time-point specific.

A further composite was also developed using a "co-primary composite" where more than one primary outcome was specified by the RCT authors. These were then compared to the original primary outcome and the targeted composite outcome generated. All analyses were performed in SPSS v24 (IBM, 2016).

5.2.3.1 **Standardised Mean Difference (SMD) Calculations**

Where data were not available for selected outcomes, then SMDs were calculated. SMDs were calculated for each primary and composite outcome for between-arm differences at the primary time-point designated by the RCT, or the soonest time-point post-exercise-intervention if no primary time-point was specified. The SMD calculation used the formula:

$$d = \frac{\dot{X}_1 - \dot{X}_2}{s} \qquad 5\text{-}1$$

Where d is SMD, Ẋ represents the mean follow-up score, 1 represents the exercise intervention group and 2 represents the control group, and s represents the average of the baseline standard deviations (106).

95% CI were calculated for the SMD, as described in chapter 3, section 3.2.6.1. SMD statistics for all between-arm differences are given based on exercise intervention minus control then positive SMDs indicate higher values for the exercise intervention (lower for the control), and by contrast, negative SMDs indicate lower values for the exercise intervention (higher for the control). Since the direction of scale data may be conflicting i.e. higher values indicate worse health outcome status (for some scales) and better health status (for other scales) – for purposes of standardisation and ease of evaluation and interpretation within the meta-analysis all SMDs were scaled such that positive SMDs reflect better outcome for the exercise intervention and negative SMDs reflect worse outcome (106). SMDs were interpreted according to Cohen's (176) recommendations, where an effect size of 0.2 or less is considered small, around 0.5 is considered medium and greater than 0.8 large.

5.2.3.2 **Secondary Analyses Performed on Obtained Datasets**

5.2.3.2.1 ***Data Modification***

As for data modification for the analyses described in chapter 4, data from obtained RCT datasets were modified in a step-wise approach as follows:

i. Datasets were checked for missing data by using descriptive analysis to check the range of inputs per matched outcome and transformed to account for excessive range where necessary (e.g. where missing data was coded with 999 or 768 these values were recoded as system missing).

ii. The data were transformed from wide to long (this was not necessary for the ANCOVA and MANCOVA) by transforming the variables to cases, and computing a new variable consisting of all time-points relevant to that outcome: for example, “Pain” (new variable) would include Pain Baseline, Pain 6-week follow-up, Pain 12-week follow-up, and Pain 6-month follow-up. The participant ID would remain the same across these time-points, and other values such as group allocation would also remain the same.

iii. Targeted outcomes were then converted to a standardised outcome in SPSS in order for the results to be comparable across a variety of outcome measurement scales, for example an outcome “Pain” would be standardised and saved as “ZPain”.

iv. Standardised composite variables were then derived by computing a new variable of the mean of the standardised variables of interest at each time-point (175).

5.2.3.2.2 *Secondary Analyses*

i. Initial analyses aimed to replicate the published or presented data used for the primary outcome(s) and/or targeted secondary outcomes where possible to do so. The replicated analysis was then applied to the matched secondary outcome(s).

    a. In one dataset the analysis replicated was ANCOVA. This analysis used the primary time-point (12-weeks). However, the authors did not specify which post hoc corrections were used, and thus a variety of corrections was used when trying to replicate the results, including Tukey, Scheffé and Bonferroni.
    b. Authors of all four RCTs performed a linear mixed model analysis on their primary outcomes. Linear mixed model analyses include all time-points available for the relevant outcome, and therefore values for all available time-points for the matched secondary outcomes were used and reported. Models were fitted including patient-identifiers as a random-effect term and including fixed-effects terms in accordance with the trial authors' specification.

ii. Further analyses were then performed using the standardised composite variables and the identified replicated method of analysis used for the primary outcome in each RCT.

iii. Two trials specified two primary outcomes each in their analysis and results paper. Exploratory analysis was undertaken to compare the results of the first nominated primary outcome in comparison to a targeted composite outcome and the co-primary outcome composite. This analysis

was only possible on the two RCTs which nominated two co-primary outcomes (136,137). The identified replicated method of analysis was applied to this composite outcome and compared with the original primary outcome results and the targeted composite outcome's results.

5.2.3.3 **Meta-analyses Performed on Summary Results**

A summary of the results was produced in forest plots using RevMan (5.3). This was created by including the primary outcome SMD in contrast to the matched composite outcome SMD for each analysed RCT in a sub-group comparison using random effects and the generic inverse variance method to demonstrate the size and direction of the effect, although meta-analysis is not possible or meaningful from these values (Figure 5-1). The Cochrane Handbook discourages the use of forest plots when only one trial's data are represented unless the pooled function is removed due to triplication of the sample size and increased risk of type I error (177).

A further between-group difference and associated 95% CI was calculated to summarise the overall between-group difference by:

i. Generating the difference for each paired between-group difference, and the overall mean between-group difference
ii. Calculating the SD and SE of the mean between-group difference
iii. Calculating the upper and lower CI limits

5.3 Results

5.3.1 Included RCTs

Of the seven ‘matched’ RCTs identified through the systematic review earlier in this thesis (120,125,131,133,135–137), only authors of two trials (136,137) responded to email requests to share their datasets. Of the five ‘unmatched’ RCTs (117,127,130,140,142), only two contained sufficient outcome domains that were matched to the exercise treatment targets for a composite to be generated (117,127). Four RCT datasets were therefore included in this analysis (n=864) (117,127,136,137). Data sharing agreements for the additional two datasets included in this analysis are included in Appendix 9.h:Data Sharing Agreement for Datasets used in Chapter 5.

Two of the included RCTs (136,137) had a primary outcome that was considered ‘matched’ in the systematic review (chapter 3) but identified more than one exercise treatment target and measured more than one matched outcome (as well as more than one primary outcome). The other two included RCTs (117,127) were categorised as ‘unmatched’, as their primary outcome was not matched to the treatment targets of their exercise intervention, but they included matched outcomes as secondary outcomes. A summary of the data extraction of the four RCTs is provided below in Table 5-1. Two RCTs used a primary time-point of 12-weeks, and the other two RCTs used a primary time-point of 6-weeks. Physical function was specified as the primary outcome in three of the four RCTs (127,136,137), although in two of the RCTs this was specified alongside another primary outcome: fear-avoidance beliefs in Moffett et al. (137) and pain in

Miyamoto et al. (136). Pain was specified as the primary outcome in two RCTs (117,136). The average number of outcomes that matched the specified exercise treatment targets was 5.5 per RCT (range 4 (127) to 7 (137)). A summary table of outcomes used in the included RCTs is given in Table 5-2.

Table 5-1: Reported treatment targets and matched outcomes of included RCTs

	Trial	Treatment Targets	Matched Outcomes		Primary Time-point	Analysis Performed	
			Primary Outcome	Secondary Outcome		Primary Outcome	Secondary Outcome
Matched	Moffett et al. (137)	Fear of physical activity Relieve pain Reduce anxiety and depression Help them take control of their situation. Enable the individual to cope better Return to their normal activities sooner Prevent long-term disability	Fear-avoidance beliefs (TSK), Physical function (RMDQ)	Health control (Multidimensional health locus of control), Self-efficacy (PSEQ), Anxiety and Depression (HADS)	6-weeks*	Each of the outcome measures (at 6-weeks, 6- and 12-months) were analysed jointly, assuming no structure for the matrix of correlations for outcomes at the three time-points. Baseline scores were adjusted by including them as covariates. The analyses for each outcome used all available data without imputation of missing data. The assumption that the missing data were missing at random was made. An initial model containing a three-way interaction between treatment allocation (McKenzie or Solution Finding), booklet allocation (received or not) and time (a factor with three levels) was fitted using maximum likelihood. Terms representing a main effect of treatment preference (a factor with three levels: McKenzie, Solution Finding or None), and its interaction with main treatment allocation were added to the simplest models permitted and tested for statistical significance. These analyses were performed using the Mixed Models procedure on SPSS Version 12 for Windows.	
	Miyamoto et al. (136)	Improving disability Reducing absence from work Physical and functional recovery Reduce pain Improve catastrophising	Pain (NRS), Physical function (RMDQ)	Physical Function (PSFS), Global Perceived Effect, Catastrophising (PCS), Fear-avoidance beliefs (TSK), Health-related Quality of Life (HRQoL) (SF6D)	6-weeks	Baseline characteristics were compared between all Pilates groups and the booklet group. The mean effects of the interventions and the group differences for all outcomes were calculated using linear mixed models that incorporate terms for the treatment groups, time (follow-ups) and interaction terms 'treatment groups' versus 'time.' The term 'time' was coded as a categorical variable (i.e., four variables were created for the categories baseline, 6-week, 6-month and 12-month follow-ups). The coefficients of treatment versus time interactions were equivalent to	

		and kinesiophobia				the estimates of the group differences. The analyses followed the intention-to-treat principle.	
Unmatched	Bronfort et al. (117)	Increase trunk muscle endurance and trunk stability	[Pain]	Static endurance (flexion, extension), dynamic endurance (flexion, extension), isometric strength (flexion, extension)	12-weeks*	Analysis of covariance (ANCOVA) was used to analyse for differences between the three groups in all patient-rated outcomes (pain was the primary outcome) at Weeks 4, 12, 26, and 52 post-randomisation. Baseline values were used as covariates. Linear mixed-model longitudinal analyses (which accounted for correlation over time within participants) were also used using the MIXED procedure in SAS 9.1.	Change scores for trunk performance measures were calculated using end treatment (Week 12) and baseline values. These were then analysed for group differences with ANOVA.
	Groessl et al. (127)	Increased strength Increased flexibility Stress reduction Increased pain tolerance.	[Physical Function]	Pain (Brief Pain Inventory), Range of motion (flexion and extension range), Core strength at 12 weeks (elbow plank time)	12-weeks	Linear mixed-effects modelling was used to examine the change score across measured time points. A main effect of group (yoga versus waiting list), a main effect of time (categorically coded for baseline, 6 weeks, 12 weeks, and 6 months) and an interaction between group X time were included in the model."	

*Bronfort et al. (115) and Moffett et al. (135) did not specify their primary time point thus the first time point post-treatment was used, as per the method used in the systematic review. Where TSK is Tampa Scale of Kinesiophobia, NRS is Numeric Rating Scale, RMDQ is Roland Morris Disability Questionnaire, PSEQ is Pain Self-Efficacy Questionnaire, HADS is Hospital Anxiety and Depression Score, NRS is Numeric Rating Scale, PCS is Pain Catastrophising Score, PSFS is Patient-Specific Functional Scale, SF-6D is the Self-Report 6 Dimension scale; ANCOVA is analysis of covariance and ANOVA is analysis of variance. Text in [] denotes unmatched primary outcomes.

Table 5-2: The outcome domains and measures used by included RCTs

	Domain Measured	Outcome Measure/ Tool	Range	Direction of Effect	Interpretation
Patient-reported Outcome Measures	Physical Function	Roland Morris Disability Questionnaire (RMDQ)(136,137)	0-24	→	0 “no disability” and 24 “maximum disability”.
		Patient-Specific Functional Scale (PSFS) (136)	0-10	←	Identify three important activities, mark on scale how capable they feel to perform these activities. 0 ‘unable to perform the activity’ and 10 ‘able to perform the activity at preinjury level’. The average of the three scores is calculated. A higher score indicates greater functional ability.
	Pain	Numeric Rating Scale (NRS) (136)	0-10	→	0 “no pain” to 10 “worst possible pain” over the past 7 days. The numerical responses are on a discrete/integer scale.
		Ordinal 11-point Box scale (117)	0-10	→	Pain over the past week, with 0 being “no pain” and 10 being “worst pain possible”
		Brief Pain Inventory (BPI) (127)	0-10	→	Pain severity subscale: Four 0-10-point numeric rating scales: rating pain “at its worst” and “at its least in the last 24 hours” and the other two asking about pain “on average” and “right now”. For each NRS, the verbal descriptors are “no pain” and “pain as bad as you can imagine”. Seven other questions relating to the pain interference subscale.
	Fear-Avoidance Beliefs	Tampa Scale of Kinesiophobia (136,137)	17-68	→	The higher the score the more severe the kinesiophobia.
	Catastrophising	Pain Catastrophising Scale (136)	0-52	→	Three subscales: rumination, helplessness, and magnification. Higher scores indicate higher catastrophising.
	Effect	Global Perceived Effect (136)	-5 to +5	←	Scores vary from ‘vastly worse’ to ‘completely recovered’. Higher scores indicate better recovery.
	Health control	Multidimensional health locus of control (137) • Internal;		←	The Internal subscale represents the degree to which a person believes he or she is in control of his or her own health. The external subscale of chance represents the degree to which

		• Powerful Others Chance.		→ →	one views fate and luck as controlling one's health outcomes, and the external subscale of the powerful others indicates the extent to which a person perceives that others control his or her health.
	Self-efficacy	Pain Self-Efficacy Questionnaire (137)	0-60	←	Higher scores indicate higher self-efficacy beliefs.
	Anxiety and Depression	Hospital Anxiety and Depression Scale (137)	0-21 0-21	→	Seven items relate to anxiety, seven items to depression. The higher the score the worse the anxiety or depression, and scores less than 7 are classed as 'non-cases'.
Objective Measures	Range of motion	Saunders digital inclinometer (127)	Measured in degrees	←	Greater range, greater function.
	Core strength	Prone and supine bridge positions (127)	Time maintained in pose to a maximum of 90 seconds	←	Greater time the better the function
	Static endurance (flexion, extension)	Biering-Sørensen test (117)	Length of time able to maintain the pose	←	Greater time the better the function
	Dynamic endurance (flexion, extension),		Maximum number of repetitions	←	Higher number of repetitions, the better the function
	Isometric strength (flexion, extension).	Computerised digital myograph (DM2000) (117)	Unclear	←	Unclear

5.3.2 Data Analysis

5.3.2.1 **Secondary Analysis of the Miyamoto et al. (2018) dataset**

The Miyamoto et al. RCT (112) compared four groups: Pilates once a week (Pilates 1), twice a week (Pilates 2), three times a week (Pilates 3) and a brief intervention without any exercise (control group). The groups were assessed at four different time-points (baseline, 6-weeks, 26-weeks, and 52-weeks) and there were six variables of interest (pain, physical function, catastrophising, fear-avoidance beliefs, global perceived effect, and a patient-specific functional scale). The first mentioned primary outcome was pain (matched), and physical function was also included as a co-primary outcome. The primary follow-up time-point was 6-weeks.

5.3.2.1.1 ***Data Modification***

The patient-specific functional scale, fear-avoidance and catastrophising scales all demonstrated lowest values of -999 in the descriptive analyses, indicating the possibility of missing data. These values were modified to reflect ‘system missing’ as any value less than zero. Four scales were multiplied by minus one (pain, physical function, catastrophising and fear-avoidance beliefs) so that all scales scored in the direction of a greater value reflecting a better outcome. The dataset was transformed from wide to long, and a standardised composite variable was created consisting of pain, physical function, catastrophising, fear-avoidance beliefs, global perceived effect, and the patient-specific functional scale.

5.3.2.1.2 *Replicating the Original Results: Linear Mixed Model Analysis of Pain*

The replicated analysis of the primary evaluation gave very similar results to those in the original RCT publication (as shown in Table 5-3 below, where the primary outcome is shaded in grey).

Table 5-3: The original RCT results and replicated results of the primary outcome (pain) in Miyamoto et al. (2018)

	Adjusted mean difference (95% CI)							
	Pilates 1 vs CG	**Sig.**	**Pilates 2 vs CG**	**Sig.**	**Pilates 3 vs CG**	**Sig**	**Int vs CG**	**Sig**
Pain (original trial results)								
6-week follow-up	1.2 (0.3, 2.2)*	<0.01	2.3 (1.4, 3.2)*	<0.001	2.1 (1.1, 3.0)*	<0.001	N/A	N/A
Secondary Analysis of Pain								
6-week follow-up	1.24 (0.34, 2.14)	0.007	2.31 1.42, 3.20)	<0.0001	2.10 (1.19, 3.01)	<0.001	1.89 (1.15, 2.63)	<0.0001

CG is control group; Pilates 1 is Pilates once a week, Pilates 2 is Pilates twice weekly, Pilates 3 is Pilates thrice weekly, CI is confidence interval, Int is all three Pilates exercise intervention groups combined. All values are calculated as the mean (intervention) minus the mean (control) values, where positive values favour the intervention arm. Shaded values reflect the primary outcome. *values reported in the RCT paper were not presented to 2 decimal points.

5.3.2.1.3 *Linear Mixed Model of a Targeted Composite Outcome*

As shown in Table 5-4, the composite produced SMDs that were smaller than the standardised pain outcome at all time-points apart from Pilates 1 compared to the control arm at 6-weeks (0.43 vs 0.44). However, the between-arm statistical significance was smaller for each of the comparisons at 6-weeks in favour of the composite outcome. Please see Appendix 9.i: Linear mixed model results of Miyamoto et al. (2018) trial dataset: all time-points.

Table 5-4: Comparison of composite outcome with primary outcome (pain) in Miyamoto et al. (2018)

	Pilates 1 vs CG			**Pilates 2 vs CG**			**Pilates 3 vs CG**			**Intervention (ALL) vs CG**	
	Effect Estimate	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig**
Standardised Pain (PRIMARY)*											
6-week follow-up	0.43 (0.04, 0.82)	0.029φ	2.193	0.88 (0.50, 1.27)	<0.0001	4.503	0.74 (0.35, 1.13)	<0.001	3.736	0.69 (0.37, 1.01)	<0.0001
Standardised Composite Outcome											
6-week follow-up	0.44 (0.21, 0.66)	**<0.001**	**3.806**	0.77 (1.00, 0.55)	<**0.0001**	**6.811**	0.59 (0.36, 0.81)	**<0.0001**	**5.079**	0.59 (0.41, 0.78)	<0.0001
Individual target-related outcomes included within the composite outcome:											
Standardised Physical Function*											
6-week follow-up	0.29 (-0.04, 0.61)	0.081	1.749	0.82 (0.50,1.14)	p<0.0001	5.032	0.53 (0.20,0.85)	0.001	3.206	0.55 (0.27, 0.81)	<0.001
Standardised Catastrophising*											
6-week follow-up	0.23 (-0.09, 0.54)	0.155	1.426	0.49 (-0.18, 0.81)	0.002	3.127	0.18 (-0.13, 0.50)	0.261	1.127	0.30 (0.05, 0.56)	0.021
Standardised Fear-avoidance beliefs*											
6-week follow-up	0.40 (0.09, 0.72)	0.012	2.524	0.50 (0.19, 0.81)	0.002	3.159	0.44 (0.12, 0.75)	0.007	2.727	0.45 (0.19, 0.70)	0.001
Standardised GPE											
6-week follow-up	0.77 (0.35, 1.19)	<0.001	3.615	1.23 (0.81, 1.64)	<0.00001	5.769	1.11(0.69, 1.54)	<0.00001	5.184	1.04 (0.69, 1.38)	<0.0001
Standardised PSFS											
6-week follow-up	0.51(0.12, 0.91)	0.01	2.58	0.72 (0.33, 1.11)	0.0003	3.651	0.54 (0.14, 0.93)	0.01	2.58	0.59 (0.27, 0.91)	0.0003

*****represent scales multiplied by minus 1. Bold items indicate where the composite value is greater than that of the single outcome CG is control group; Pilates 1 is Pilates once a week, Pilates 2 is Pilates twice weekly, Pilates 3 is Pilates thrice weekly, Intervention ALL is the combined effect of all three Pilates exercise intervention arms, CI is confidence interval. GPE is global perceived effect, PSFS is Patient-Specific Functional Scale; φ This P-value does not match exactly with the 0.007 in Table 3 possibly due to the slightly different correlation structures that are modelled in (perhaps these may also have been slightly different in specification i.e. unstructured vs autoregressive). Further detail regarding the other reported time-points can be found in Appendix 1

Further, the standard error of the composite outcome was much smaller than that of the standardised pain outcome, as seen in Table 5-5.

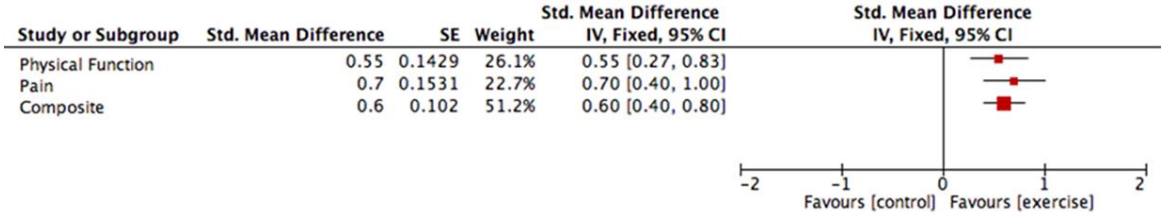
Table 5-5: The standard error of the two outcome variables in Miyamoto et al. (2018)

Parameter	Standardised Pain Outcome	Standardised Composite Outcome
FU1 Pilates 1 vs CG	0.08	0.04
FU1 Pilates 2 vs CG	0.06	0.03
FU1 Pilates 3 vs CG	0.09	0.05

CG is control group; FU is the follow-up time-point where FU1 is 6-weeks

In the included components of the composite outcome, all the individual measures demonstrated a between-arm statistically significant difference in favour of exercise in the Pilates 2 group at 6-weeks. The standardised catastrophising outcome was the only measure not to demonstrate a statistically significant difference at 6-weeks between the Pilates 1 and Pilates 3 times-a-week groups in comparison to the control arm, and the physical function outcome was not statistically significantly different in the Pilates 1 at 6-weeks in comparison to the control arm. When all the Pilates exercise intervention arms were combined into one group and compared to the control arm, there was a statistically significant between-arm difference at the 6-week follow-up time-point in favour of exercise, with the composite outcome in comparison to the original primary outcome of pain. Further, the SMD was similar between the standardised pain score and composite outcome score at the primary time-points, although standardised pain at the first time-point (6-weeks) was 0.1 points greater than the composite outcome. The results are summarised in Figure 5-1.

Figure 5-1: *The size and direction of effects of the primary and composite outcomes in Miyamoto et al. (2018)*



Where SE represents the standard error, IV inverse variance, CI is confidence interval, Std is standardised as part of SMD

5.3.2.1.4 *Linear Mixed Model Analysis of the Co-Primary Composite*

The linear mixed model was used for the analysis, as described in 5.2.3.2. The two nominated primary outcomes were pain and physical function. The results are summarised below in Table 5-6. This suggests that the co-primary composite outcome may perform as well as the single matched primary outcome, in this trial, with little loss of SMD size, and increase in statistical significance. Further, the co-primary composite demonstrated greater SMD than the composite in three of the intervention arms in comparison to the control arm, with similar statistical significance to the targeted composite outcome, with only one time-point and group generating a statistical significance that was larger than the composite (but smaller than the primary outcome).

Table 5-6: A comparison of the primary outcome (pain) to a targeted composite and co-primary outcome composite in Miyamoto et al. (2018)

	Pilates 1 vs CG			**Pilates 2 vs CG**			**Pilates 3 vs CG**			**Intervention All vs CG**	
	Effect Estimate	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**
Standardised Pain (PRIMARY)											
6-week follow-up	0.43 (0.04, 0.82)	0.029φ	2.193	0.88 (0.50, 1.27)	<0.0001	4.503	0.74 (0.35, 1.13)	<0.001	3.736	0.69 (0.37, 1.01)	<0.0001
Standardised Composite Outcome											
6-week follow-up	0.44 (0.21, 0.66)	**<0.001**	**3.806**	0.77 (1.00, 0.55)	<**0.0001**	**6.811**	0.59 (0.36, 0.81)	**<0.0001**	**5.079**	0.59 (0.41, 0.78)	<0.0001
Standardised Co-Primary Composite Outcome											
6-week follow-up	0.36 (0.06, 0.65)	**0.017**	**2.400**	0.85 (0.56, 1.14)	<0.0001	**5.723**	0.64 (0.33, 0.93)	<0.0001	**4.215**	**0.62 (0.37, 0.86)**	**<0.0001**
Individual target-related outcomes included within the composite outcome:											
Standardised Physical Function											
6-week follow-up	0.29 (-0.04, 0.61)	0.081	1.749	0.82 (0.50,1.14)	p<0.0001	5.032	0.53 (0.20,0.85)	0.001	3.206	0.55 (0.27, 0.81)	<0.001

Bold items indicate where the composite value is greater than that of the single outcome. CG is control group; Pilates 1 is Pilates once a week, Pilates 2 is Pilates twice weekly, Pilates 3 is Pilates thrice weekly, Intervention ALL is the combined effect of all three Pilates exercise intervention arms CI is confidence interval. GPE is global perceived effect, PSFS is Patient-Specific Functional Scale

5.3.2.1.5 *Summary of Miyamoto et al. (2018)*

The composite outcome variable produced between-arm estimates that are more statistically significantly different (Pilates 1 t=3.806, p<0.0001; Pilates 2, t=-6.811, p<0.0001; Pilates 3 t=-5.079, p<0.0001) than the original standardised pain primary outcome (Pilates 1 t=-2.193, p=0.029; Pilates 2, t=-4.503, p<0.0001; Pilates 3 t=-3.736, p<0.0001) with a smaller standard error than the original standardised variable (composite 0.04, pain 0.08). When all Pilates exercise intervention groups were combined, the composite outcome produced a between-arm estimate effect that was comparable to that achieved using the original primary outcome and was also statistically significant at the primary time-point. However, the co-primary composite outcome generated a slightly higher SMD than the targeted composite outcome for Pilates 2, Pilates 3 and all intervention arms combined, although these values were still slightly smaller than the matched primary outcome in all arms at the primary time-point. It appears from the results of the secondary analysis of this first RCT (136) that a single, matched outcome appeared to generate greater SMDs and statistical significance, than a co-primary and targeted composite of multiple exercise treatment targets.

5.3.2.2 **Secondary Analysis of Moffett et al. (2006) Dataset**

The Moffett et al. RCT (137) compared a McKenzie exercise approach with a 'solution-finding' approach (SFA) (comprising goal setting using cognitive behavioural methods). McKenzie is a specific method that has a strong emphasis on exercise but can contain some manual therapy (178). This RCT had four outcome assessment time-points (baseline, 6-weeks, 24-weeks, and 52-weeks) and six variables of interest (fear-avoidance beliefs, physical function, health control, self-efficacy, anxiety and depression). The first mentioned primary outcome was fear-avoidance beliefs (matched), and the second nominated primary outcome (physical function) was included in a co-primary composite outcome. The authors did not specify their primary time-point, so this was extracted as per the process described in chapter 3, section 3.2.6.1, as the first follow-up time-point at 6-weeks.

5.3.2.2.1 ***Data Modification***

The descriptive statistics of all six variables of interest at all four time-points were checked, and no missing data were identified. The dataset was transformed from wide to long as described in 5.2.3.2.1. The subscales of health control (powerful others and chance), anxiety and depression, physical function and the combined fear-avoidance beliefs scales were multiplied by minus one, and a standardised composite outcome was developed comprising of fear-avoidance beliefs, physical function, health control (including subscales of chance, internal and external), self-efficacy, anxiety and depression.

5.3.2.2.2 *Testing the Original Results: Linear Mixed Model Analysis of Pain*

The analysis on the primary outcome (fear-avoidance beliefs) was replicated and produced results that were similar to the reported figures in the RCT paper, as seen in Table 5-7 below. At only one of the time-points were the between-arm differences statistically significant (24-weeks), and this was replicated with a p-value of 0.012.

Table 5-7: Replication of the trial results of Moffett et al. (2006)

	Adjusted mean difference (95% CI)	
McKenzie vs SFA	**Effect Estimate**	**Sig.**
Fear-Avoidance Beliefs (trial results) (TSK-AA)		
6-week follow-up	-0.04 (-0.98, 0.90)	0.032
24-week follow-up	1.12 (0.15, 2.09)	
52-week follow-up	0.39 (-0.62,1.40)	
This Analysis of Fear-Avoidance Beliefs (TSK-AA)		
6-week follow-up	-0.04 (-0.99, 0.90)	0.928
24-week follow-up	1.16 (0.26, 2.06)	0.012
52-week follow-up	0.43 (-0.52, 1.38)	0.376

Where SFA represents Solution Finding Approach; TSK represents the Tampa Scale of Kinesiophobia Activity Avoidance subscale.

5.3.2.2.3 *Linear Mixed Model of a Targeted Composite Outcome*

The model was constructed using a composite outcome variable composed of the six matched outcomes described earlier. As per the original analysis, the baseline variables and booklet allocation were controlled for as covariates. Table 5-8 demonstrates the results of the composite outcome variable in comparison to the standardised fear-avoidance beliefs variable. Although the original primary outcome results demonstrated a statistically significant difference between the groups at 24-weeks, the composite variable did not demonstrate a greater difference nor statistical significance. Please see Appendix 9.j: Linear mixed model results of Moffett et al. (2006) trial dataset: all time-points.

Table 5-8: The results of the standardised composite outcome in comparison to the standardised primary outcome measure for Moffett et al. (2006)

	Adjusted mean difference (95% CI)		
	McKenzie vs SFA		
Standardised Fear-Avoidance Beliefs (TSK-AA) *	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	-0.01 (-0.22,0.20)	0.94	-0.079
Standardised Composite Outcome	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	-0.01 (-0.11,0.09)	0.868	-0.167
Individual Components of the Composite			
Standardised fear-avoidance beliefs (TSK-SF and TSK-AA combined) *			
6-week follow-up	-0.11 (-0.28, 0.06)	0.214	-1.25
Standardised physical function *			
6-week follow-up	0.14 (-0.09,0.37)	0.227	1.21
Standardised health control (internal)			
6-week follow-up	-0.03 (-0.16, 0.22)	0.763	0.302
Standardised health control (chance scale) *			
6-week follow-up	0.08 (-0.11,0.27)	0.413	-0.820
Standardised health control (powerful others scale) *			
6-week follow-up	-0.07 (-0.25,0.12)	0.491	-0.690
Standardised self-efficacy			
6-week follow-up	0.01 (-0.20, 0.18)	0.925	0.094
Standardised anxiety *			
6-week follow-up	0.08 (-0.09, 0.24)	0.357	0.92
Standardised depression*			
6-week follow-up	0.05 (-0.13, 0.22)	0.590	0.540

Bold italics items indicate where statistical significance; Where SFA represents Solution Finding Approach and TSK represents the Tampa Scale of Kinesiophobia; TSK-AA is the Activity Avoidance subscale and the TSK-SF is the somatic focus subscale. Asterisks denote the scale has been multiplied by minus one to trend in a positive direction. Data for all other time points can be found in Appendix 2.

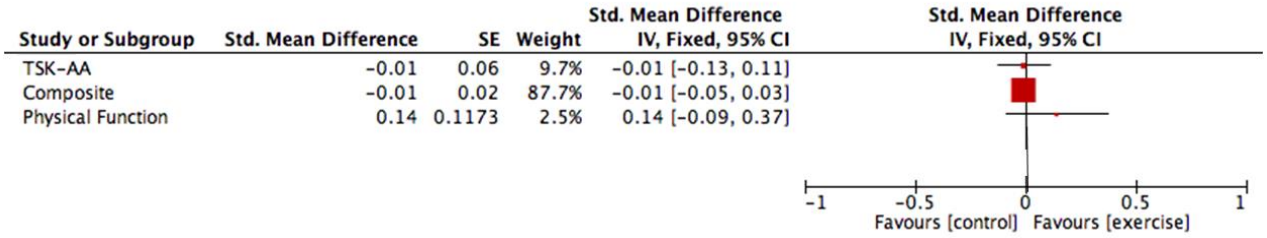
Table 5-9 demonstrates a smaller standard error in the composite outcome variable compared to the RCT's primary outcome variable.

Table 5-9: The standard error of the two outcome variables in Moffett et al. (2006)

Parameter	Standardised Fear-Avoidance Outcome	Standardised Composite Outcome
1,1	0.06	0.02
2,1	0.06	0.02
2,2	0.07	0.02

Figure 5-2 demonstrates the SMDs of the primary outcomes and composite outcome as seen in Table 5-8 in a forest plot, for comparison of effect size and direction at the primary time-point.

Figure 5-2: The effect of primary and composite outcomes in Moffett et al. (2006)



Std. is standard as part of SMD, SE is standard error, CI is confidence interval, IV is inverse variance

5.3.2.2.4 *Linear Mixed Model Analysis of the Co-Primary Composite*

The linear mixed model was used for the analysis, as described in section 5.3.2.2. The two nominated primary outcomes were fear-avoidance beliefs (Tampa Scale of Kinesiophobia activity-avoidance subscale) and physical function (Roland and Morris Disability Questionnaire). The results are tabulated below in Table 5-10. They suggest that the co-primary composite outcome generated a greater effect estimate than the targeted composite outcome and the single matched outcome, with between-arm SMDs that were greater than the targeted composite at the primary time-point.

Table 5-10: Table to demonstrate the comparison of the primary outcome (fear-avoidance beliefs) in comparison to a targeted composite and co-primary outcome composite in Moffett et al. (2006)

	Adjusted mean difference (95% CI)		
	McKenzie vs SFA		
Standardised Fear-Avoidance Beliefs (TSK-AA)	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	-0.01 (-0.22,0.20)	0.94	-0.079
Standardised Targeted Composite Outcome	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	0.00 (-0.08,0.08)	0.89	0.006
Standardised Co-Primary Composite			
6-week follow-up	0.08 (-0.13,0.29)	0.45	0.756
Standardised Physical Function			
6-week follow-up	0.14 (-0.09,0.37)	0.227	1.21

Where TSK-AA represents Tampa Scale of Kinesiophobia activity-avoidance subscale; SFA is solution-finding approach.

5.3.2.2.5 *Summary*

This analysis of a second RCT, of four in this study, demonstrates that a matched primary outcome measure produced the same between-arm effect estimate at the first time-point, as a matched composite outcome measure. However, the co-primary matched composite demonstrated greater effect estimates at the primary time-point. Most of the matched outcomes comprising the composite outcome (fear-avoidance beliefs, physical function, anxiety and depression, and the subscales of health control) demonstrated greater SMDs than the single primary outcome at six-weeks (albeit two of the health control scales and the fear-avoidance scales favoured the control group), but these were not maintained at other time-points. The SMD was greatest in the co-primary composite at the primary time-point (6-weeks) (SMD 0.08 (95% CI -0.13, 0.29)) in contrast to the single matched primary outcome (SMD -0.01 (95% CI -0.22, 0.20)) and the

targeted composite (SMD 0.00 (95% CI -0.08, 0.08)), but none of these results were statistically significant.

5.3.3 Secondary Analyses of Unmatched Datasets

5.3.3.1 **Secondary Analysis of Bronfort et al. (2011)**

The Bronfort et al. RCT (115) had four time-points (baseline, 12-weeks, 26-weeks, and 52-weeks) and six variables of interest (dynamic endurance flexion and extension strength, static endurance flexion and extension strength, isometric flexion and extension strength). However, these variables were only measured at baseline and 12-weeks. They compared specific exercise therapy (SET) with spinal manual therapy (SMT) and home exercise therapy (HEA) control. Their primary outcome was pain, and they did not report any statistically significant between-arm differences in the primary outcome at any of the time-points. The primary analysis method was ANCOVA with baseline variables as covariates, although the linear mixed model analysis results were reported in the published paper.

5.3.3.1.1 ***Standardised Mean Differences of the Original Analysis Results***

The pain variable was not requested in the dataset as it was not a matched outcome; thus, it was not possible to replicate the original analysis on the primary outcome. The SMD of each comparison was calculated for each time-point using the methods described in section 5.2.3. The results for the primary time-point are displayed in Table 5-11. The reported between-arm differences were not

statistically significant with all CIs including zero (using a general linear mixed model).

Table 5-11: Published trial results of the primary outcome measure (pain) in Bronfort et al. (2011)

PAIN	Wk. 12 (Primary)	SMD
SET* vs SMT	0.1 (-0.6, 0.8)	*0.21 (-0.07,0.5)*
SET* vs HEA	0.6 (-0.1, 1.4)	*0.43 (0.14, 0.72)*
SMT* vs HEA	0.6 (-0.2, 1.3)	*0.20 (-0.08, 0.48)*

Where HEA represents home exercise and advice, SMT represents spinal manual therapy, SET represents specific exercise therapy and SMD represents standardised mean difference; Negative values favour the control group (denoted with an asterisk, where the intervention is the active comparator e.g. mean score (SET) – mean score (SMT) for the first data row); italicised data are calculated SMD values; italicised data reflects calculated SMD values

5.3.3.1.2 *Data modification*

Of the twelve matched outcome variables (six matched outcomes at two time-points: baseline and 12-week follow-up), there were no missing data and all variable scales scored in the same direction. For the linear mixed model to be run, the dataset was transformed from wide to long, by merging the twelve original variables into six standardised variables as described in section 5.2.3.2.1.

5.3.3.1.3 *Replication of the ANCOVA Analyses*

Replication of the original analyses was performed in chapter 4 (section 4.3.2.2.1.2) for the secondary outcomes matched to the treatment targets of the exercise intervention. Further analysis of the composite outcome was performed as per the method used for the primary outcome analysis. MANCOVA was also performed to compare the results with the ANCOVA of the standardised composite. ANCOVA analyses were used to compare the differences between

three groups at all time-points with baseline values as covariates – this was conducted in two separate analyses to allow for the differing reference group (i.e. HEA or SMT) (please see chapter 4, section 4.3.2.2.1.2). The results are tabulated in Table 5-12. In comparison to the standardised reported pain scores, the composite outcome values provided a greater SMD in comparison to manual therapy (0.21 vs 0.26), with smaller confidence intervals, suggesting a more precise effect estimate that is statistically significant ($p<0.00001$). The composite outcome was also statistically significant in favour of exercise in comparison to home exercise and advice at 12-weeks although the original primary outcome generated a greater SMD (0.43 vs 0.29), as seen in Table 5-12. The components comprising the composite outcome (isometric extension strength, static endurance flexion and extension, dynamic endurance flexion and extension) were all statistically significant in favour of the exercise arm apart from standardised isometric flexion strength outcome (in comparison to the home exercise arm).

Table 5-12: ANCOVA analysis of composite outcome in comparison to the primary outcome (pain) in Bronfort et al. (2011)

Outcome*/ Group Comparison	Mean difference (95% Confidence Intervals)	Significance (<0.05)	Standard Error
Standardised Primary (Pain) Outcome			
SET vs SMT	0.21 (-0.07, 0.5)	0.156	0.148
SET vs HEA	0.43 (0.14, 0.72)	0.004	0.148
Standardised Composite			
SET vs SMT	0.26 (0.16,0.36)	<0.0001	0.051
SET vs HEA	0.29 (0.19, 0.39)	<0.0001	0.053
Standardised Variables comprising the Composite Outcome:			
Dynamic Endurance Flexion			
SET vs SMT	0.59 (0.39, 0.79)	<0.0001	0.101
SET vs HEA	0.65 (0.44, 0.86)	<0.0001	0.106
Dynamic Endurance Extension			
SET vs SMT	0.84 (0.71,1.11)	<0.0001	0.097
SET vs HEA	0.91 (0.71, 1.11)	<0.0001	0.101
Static Endurance Flexion			
SET vs SMT	0.57 (0.35, 0.78)	<0.0001	0.108
SET vs HEA	0.44 (0.22,0.67)	<0.0001	0.113
Static Endurance Extension			
SET VS SMT	0.32 (0.13,0.52)	0.001	0.100
SET VS HEA	0.40 (0.19,0.60)	0.0002	0.105
Isometric Flexion Strength			
SET VS SMT	0.20 (0.05,0.35)	0.011	0.076
SET VS HEA	-0.01 (-0.16,0.15)	0.929	0.079
Isometric Extension Strength			
SET VS SMT	0.19 (0.04,0.33)	0.015	0.075
SET VS HEA	0.17 (0.01,0.32)	0.034	0.078

All outcome values are for standardised outcomes.

5.3.3.1.4 *MANCOVA Analysis*

MANCOVA is an alternative method of analysing multivariate outcomes and was used as a secondary multivariate analysis in the trial paper. Thus, the data were also analysed in this study using MANCOVA to compare to the composite ANCOVA analysis above (see Table 5-13). This demonstrated that the combined results of the outcomes of interest were statistically significant across all four tests, and all standardised components, although no effect estimate can be derived.

Table 5-13: Table to show MANCOVA of standardised outcomes comparing exercise against manual therapy and home exercise in Bronfort et al. (2011)

Outcome	**F Value**	**Partial ETA Squared**	**Sig.**	**Outcome**	**F Value**	**Partial ETA Squared**	**Sig.**
Group (SET=1 and SMT=0) (Composite)				**Group (SET=1 and HEA=2) Composite**			
Pillai's Trace	.344	.344	P<0.000001	Pillai's Trace	0.318	0.318	P<0.000001
Wilk' Lambda	.656	.344	P<0.000001	Wilk' Lambda	0.682	0.318	P<0.000001
Hotelling's Trace	.525	.344	P<0.000001	Hotelling's Trace	0.466	0.318	P<0.000001
Roy's Largest Root	.525	.344	P<0.000001	Roy's Largest Root	0.466	0.318	P<0.000001

Sig is significance; SET is specific exercise, SMT is spinal manual therapy, and HEA is home exercise.

5.3.3.1.5 *Replication of the Linear Mixed Model Analyses*

A linear mixed model of the composite was also created to compare against the primary outcome (Table 5-14). This demonstrated that the composite outcome generated a slightly smaller SMD than the standardised pain outcome when comparing specific exercise to home exercise (although this result was statistically significant in contrast to the primary outcome). When comparing specific exercise to spinal manual therapy, there was a statistically significant and greater SMD found with the composite in favour of specific exercise (SMD 0.21 (95%CI -0.07, 0.5) for pain compared to SMD 0.43 (95% CI 0.31, 0.54) for the composite).

Table 5-14: The linear mixed model analysis of standardised outcomes comparing exercise against home exercise and manual therapy in Bronfort et al. (2011)

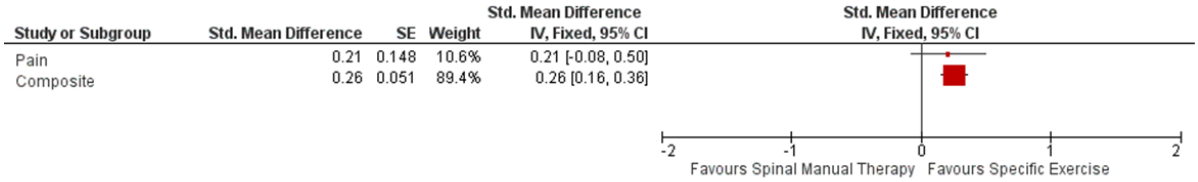
	Adjusted mean difference (95% CI)		
12-week follow-up	**Effect estimate (95% CI)**	**t score**	**Sig.**
Standardised Pain (trial results)			
SET* (vs HEA)	0.43 (0.14, 0.72)		Not reported
SET* (vs SMT)	0.21 (-0.07, 0.5)		Not reported
Composite Results			
SET* (vs HEA)	0.41 (0.29, 0.54)	6.595	<0.000001
SET* (vs SMT)	0.43 (0.31, 0.54)	7.076	<0.000001
Individual Components of the Composite Outcome:			
Standardised Static endurance Flexion			
SET* vs SMT	0.55 (0.32, 0.79)	4.606	<0.0001
SET* vs HEA	0.40 (0.16, 0.65)	3.228	0.001
Standardised Static endurance Extension			
SET* vs SMT	0.31 (0.09, 0.52)	2.815	0.005
SET* vs HEA	0.36 (0.14, 0.58)	3.174	0.002
Standardised Dynamic Endurance Flexion			
SET* vs SMT	0.56 (0.34, 0.78)	4.965	<0.00001
SET* vs HEA	0.63 (0.40, 0.86)	5.41	<0.00001
Standardised Dynamic Endurance Extension			
SET* vs SMT	0.84 (0.62, 1.05)	7.635	<0.00001
SET* vs HEA	0.92 (0.70, 1.14)	8.098	<0.00001
Standardised Isometric Strength Flexion			
SET* vs SMT	0.15 (-0.00, 0.31)	1.932	0.054
SET* vs HEA	0.003 (-0.16, 0.16)	0.033	
Standardised Isometric Strength Extension			
SET* vs SMT	0.01 (-0.15, 0.17)	0.130	
SET* vs HEA	0.18 (-0.02, 0.34)	2.208	0.028

Where HEA represents home exercise and advice, SMT represents spinal manual therapy, and SET represents specific exercise therapy.

5.3.3.1.6 *Summary of Bronfort et al. (115)*

Using this third trial's primary analysis method (ANCOVA), the composite outcome generated a greater SMD (0.26 (95% CI 0.16, 0.36), p<0.00001) in favour of exercise than the primary outcome (SMD 0.21 (95% CI -0.07, 0.5), ns), as seen in the forest plot, Figure 5-3.

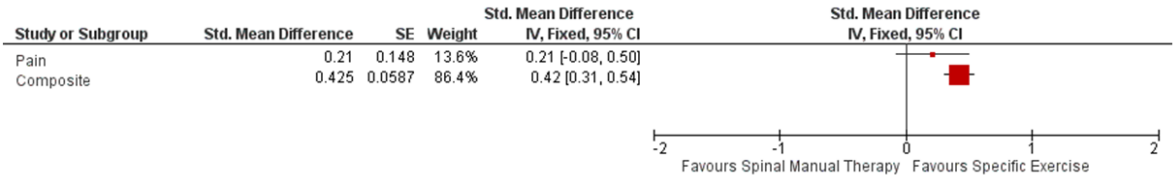
Figure 5-3: The effect of the primary outcome (pain) in comparison to the composite for exercise in comparison to manual therapy using an ANCOVA analysis in Bronfort et al. (2011)



SE is standard error, IV is inverse variance method, CI is confidence interval, Std. is standard as part of SMD. All values represent the mean difference between specific exercise and spinal manual therapy.

The results of the linear mixed model (see Figure 5-4) also demonstrated that the use of a composite variable appear to change the conclusion of the trial in favour of specific exercise therapy with a larger effect estimate compared to both home exercises and spinal manual therapy, with a SMD of 0.43 ((95% CI 0.31, 0.54), p<0.00001) in comparison to the original trial results (SMD 0.21 (95% CI -0.5, 0.07)).

Figure 5-4: The effect of the primary outcome (pain) in comparison to the composite for exercise in comparison to spinal manual therapy using a linear mixed model analysis in Bronfort et al. (2011)



SE is standard error, IV is inverse variance method, CI is confidence interval, Std. is standard as part of SMD. All values represent the mean difference between exercise and spinal manual therapy.

Further analysis of a composite derived by MANCOVA demonstrated statistically significant results for all values at 12-weeks when comparing supervised exercise to manual therapy and supervised exercise to home exercise. The linear mixed model, which is the analysis reported in the published paper, is preferential to the ANCOVA or repeated measures ANOVA or MANCOVA by making greater use of the available data within the analysis and by modelling missing data more appropriately to the covariates in the model.

5.3.3.2 **Secondary Analysis of Groessl et al. (2017)**

The last of the four RCTs analysed in this study, by Groessl et al. (126) had four time-points (baseline, 6-weeks, 12-weeks, and 24-weeks) and three secondary outcomes that matched their exercise treatment targets (strength, flexibility and pain relief). The primary outcome was physical function (unmatched) at 12-weeks and results showed statistically significant differences between yoga and waitlist control at the first two follow-ups, with statistically significant differences at 24-weeks in favour of the yoga arm.

5.3.3.2.1 ***Standardised Mean Differences of the Original Analysis Results***

The physical function variable was not requested in the dataset as it was not a matched outcome; thus, the original analysis was not able to be replicated on the primary outcome. SMD values were calculated for the reported primary outcome (physical function, shaded in grey) and compared to the only other reported matched outcome (pain) as seen in Table 5-15.

Table 5-15: Calculated SMD values comparing reported primary outcome domain and matched secondary outcomes in Groessl et al. (2017)

Trial	**Comparator**	**Outcome Domain (Primary outcome shaded)**	**Standardised Mean Difference (95% Confidence Interval)**
Groessl et al. (2017)	Yoga vs Waiting list	Physical Function	0.14 (-0.27, 0.55)
		Pain	0.30 (0.08, 0.52)

5.3.3.2.2 *Data Modification*

On reviewing the outcomes of interest (four outcomes) via descriptive variables, there were no missing data identified. Both strength and the flexion flexibility variables required multiplication by minus one so that the scales all pointed in the same direction. The dataset was converted from a wide to a long format as described in section 5.2.3.2.1 by merging all four time-points included for each targeted outcome.

5.3.3.2.3 *Linear Mixed Model of a Targeted Composite Outcome*

The linear mixed model method applied to the primary outcome was tested and described previously (chapter 4, section 4.3.2.2.2.2). The composite outcome was composed of strength, flexibility and pain outcomes. The composite outcome demonstrated effect estimates that are the same as the primary outcome at six weeks but are greater than the primary outcome at the primary time-point (12-weeks), and this was also statistically significant in favour of the exercise group (composite outcome SMD 0.25 (95% CI -0.43, -0.05) versus primary outcome SMD -0.14 (95% CI -0.46, 0.18)) as seen in Table 5-16. The individual components of the composite outcome demonstrate that for three of the four

variables, greater SMDs were observed than using the standardised primary outcome (standardised pain, plank and flexion ROM) at the primary time-point. Although, only standardised pain showed differences that were statistically significant in favour of the exercise arm. Please see Appendix 9.k: Linear mixed model results of Groessl et al. (2017) trial dataset: all time-points.

Table 5-16: The linear mixed model analysis of the standardised pain outcome in comparison to the composite outcome results in Groessl et al. (2017)

	Adjusted mean difference (95% CI)		
Yoga vs WL	**Effect estimate (95% CI)**	**t score**	**Sig.**
Standardised Primary Outcome (Physical Function) (trial results)			
12-week follow-up	-0.14 (-0.46,0.18)		0.34
Standardised Composite Analysis			
12-week follow-up	-0.25 (-0.43, -0.07)	-2.761	0.007
Individual Components of the Composite Outcome:			
Standardised Pain			
12-week follow-up	*-0.30 (-0.52, -0.08)*	*-2.741*	*0.007*
Standardised Plank			
12-week follow-up	-0.23 (-0.51, 0.04)	-1.64	0.105
Standardised Flexion ROM			
12-week follow-up	-0.27 (-0.61, 0.08)	-1.538	0.127
Standardised Extension ROM			
12-week follow-up	-0.08 (-0.44, 0.28)	-0.456	0.649

Where WL is waiting list, ROM is range of motion. All values are calculated as mean difference of yoga minus mean difference of waiting list, where negative values favour the yoga intervention. Please see appendix B3 for the analysis relating to other reported time-points. Values in italics are statistically significant.

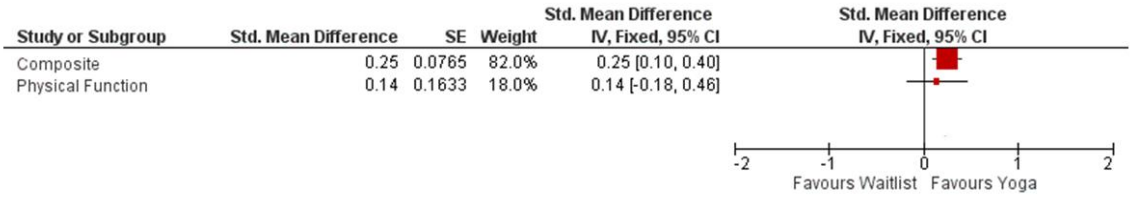
The standard error of the composite outcome was much smaller than the standardised pain outcome (0.06 vs 0.12) as seen in Table 5-17.

Table 5-17: The standard error of the two models in Groessl et al. (2017)

Parameter	Standardised Pain Outcome	Standardised Composite Outcome
1,1	0.12	0.06
2,1	0.11	0.05
2,2	0.12	0.05
3,1	0.11	0.04
3,2	0.16	0.04
3,3	0.14	0.05
4,1	0.10	0.04
4,2	0.10	0.04
4,3	0.11	0.04
4,4	0.11	0.04

Figure 5-5 demonstrates a forest plot of the primary outcome in comparison to the composite outcome at the primary time-point.

Figure 5-5: The effect of the primary outcome (physical function) in comparison to the composite in comparison to waiting list control in Groessl et al. (2017)



SE is standard error, IV is inverse variance, CI confidence intervals, Std. is standard as part of SMD

5.3.3.2.4 *Summary of Groessl et al. (126)*

The results using the standardised composite outcome measure incorporating the stated exercise targets in this fourth RCT (127) differed to those observed using the original primary outcome. A statistically significant difference was found in favour of the exercise arm at 12-weeks in the composite outcome (the primary time-point), and a greater SMD was found at the primary time-point in favour of

the exercise arm with the composite outcome. In the original analysis of this RCTs' primary outcome, a statistically significant difference between exercise versus control was only found at 24-weeks. The matched outcomes comprising the composite variable found greater SMDs at the primary time-point in comparison to the primary outcome (apart from extension ROM), but only pain was statistically significantly different at this time-point.

5.3.3.3 **Overall summary from secondary analyses of all four RCTs**

Analyses using standardised composite outcomes compared to the initial primary outcome domain in this sample of four RCTs demonstrated that in three of four trials, the use of a matched composite outcome generated a greater standardised between-arm mean difference in favour of the exercise intervention. The new analysis changed the results and conclusions of the two 'unmatched' RCTs (117,127) and one of the two 'matched' RCTs (136,137), as seen in Table 5-18. Two of the four analyses showed results with the composite outcome variable that had slightly smaller SMDs in favour of the exercise intervention, whilst the other two found greater SMDs in favour of the exercise intervention. Three of these results were (more) statistically significant in comparison to the original RCTs' primary outcome results. All analyses showed a smaller standard error when using the composite outcome. However, if one considers the matched status, then in the two RCTs with matched primary outcomes, the matched single primary outcome performed better than the composite outcome; and in the two RCTs with unmatched primary outcomes, the matched composite performed

better than the unmatched primary outcome, potentially changing the conclusion of both RCTs with an unmatched primary outcome.

Table 5-18: Summary results of all four RCTs primary outcome and targeted composite outcome

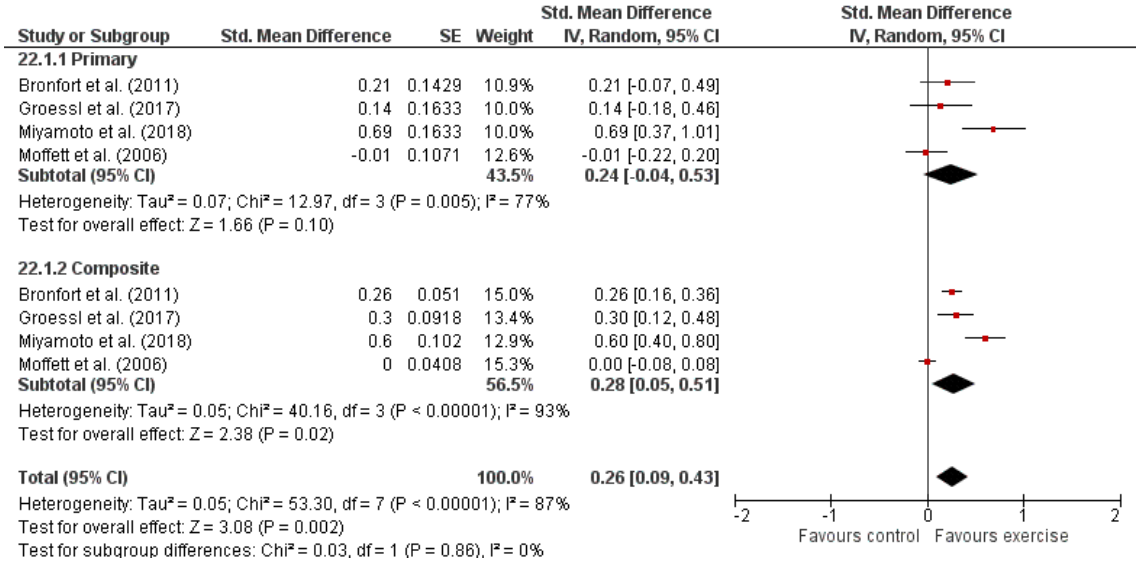
	Trial	Primary Time-Point	Outcome	SMD (Brackets denote 95% confidence intervals)	Sig. (at p<0.05)	Conclusion
Matched	Miyamoto et al. (2018)	6-weeks	Primary (Pain)	0.69 (0.4, 1.0)	<0.0001	No change
			Composite	0.60 (0.4, 0.8)	<0.0001	
	Moffett et al. (2006)	6-weeks	Primary (Fear-Avoidance Belief)	-0.01 (-0.22, 0.20)	NS	No change
			Composite	0.00 (-0.08,0.08)	NS	
Unmatched	Bronfort et al. (2011)	12-weeks	Primary (Pain)	0.21 (-0.07, 0.5)	Not reported	Changed results in favour of exercise
			Composite (ANCOVA)	0.26 (0.16,0.36)	*<0.0001*	
			Composite (LMM)	0.43 (0.31, 054)	*<0.0001*	
	Groessl et al. (2017)	12-weeks	Primary (Physical Function)	0.14 (-0.46,0.18)	NS	Changed results in favour of exercise
			Composite	0.30 (0.08, 0.52)	0.007	

Where NS is non-significant, SMD is standardised mean difference, LMM is linear mixed model, ANCOVA is analysis of variance with co-variates.

A meta-analysis of the summary results provides further support in favour of using the composite outcome given the observed pooled SMD in favour of the composite (SMD 0.28 (95%CI 0.05, 0.51) p=0.02), which was statistically significant, in contrast to the single primary outcome (SMD 0.24 (95%CI -0.04, 0.53) p=0.10) as seen in Figure 5-6, although the sub-group differences were not statistically significant (p=0.86). The between-group difference when calculated was SMD 0.03 (95% CI -0.13, 0.20) p=0.86. This difference may be because three of the composite outcomes had statistically significant results in favour of the exercise arm, in contrast to only one of the single outcomes, and the standard

errors were much smaller in the composite outcomes (range -0.041 to 0.102) in contrast to the single outcomes (range 0.107 to 0.163).

Figure 5-6: Summary plot to demonstrate pooled SMD of primary outcome in comparison to composite outcome



Std. represents standard as part of SMD, SE is standard error, IV is inverse variance, CI is confidence interval.

In the Miyomoto et al. (136) analyses, the co-primary composite generated a greater SMD than the targeted composite outcome, but this was still not as large as the single matched primary outcome. In the Moffett et al. (137) analysis, the co-primary composite generated a greater SMD than both the single matched outcome and the targeted composite outcome in favour of exercise in comparison to a solution finding approach. These results are summarised in Table 5-19.

Table 5-19: Summary table of co-primary, targeted composite and primary matched outcomes

Trial	Primary Time-point and comparison	Outcome	SMD (95% CI)	Sig. (at p<0.05)
Miyamoto et al. (136)	6-weeks (Intervention all vs CG)	Primary (Pain)	0.69 (0.4, 1.0)	<0.0001
		Composite	0.60 (0.4, 0.8)	<0.0001
		Co-primary composite	0.62 (0.37, 0.86)	<0.0001
Moffett et al. (137)	6 weeks (McK vs SFA)	Primary (Fear-Avoidance)	-0.01 (-0.22,0.20)	NS
		Composite	0.00 (-0.08,0.08)	NS
		Co-primary composite	0.08 (-0.13,0.29)	NS

SMD represents standardised mean difference, CI is confidence interval, Sig. is significance, CG is control group, McK is McKenzie method, SFA is solution finding approach

5.4 Discussion

This exploratory secondary analyses of existing RCTs developed and evaluated the use of composite primary outcomes, matched to the stated exercise treatment targets in four example RCT datasets. The results suggest that, in three of the four RCTs, use of a composite primary outcome would have potentially altered the results and conclusions reached by these trial teams. Using a composite outcome, matched to multiple exercise treatment targets, may give greater power to detect superiority of exercise interventions over a non-exercise comparator. Greater between-arm SMDs in favour of the exercise intervention, that were statistically significant compared to the estimates derived using the trials' original primary outcomes, with smaller standard errors, were demonstrated in three RCT datasets in favour of exercise in comparison to a control arm (117,127,137). Miyamoto et al. (136) was the only RCT that found SMD results greater with the original matched primary outcome. It is important to

note that in this secondary analysis of four RCTs, two included matched primary outcomes, compared to matched composites, and two RCTs included unmatched primary outcomes compared to matched composites. Despite this, the overall pooled mean difference still favoured the use of a matched composite to a single outcome (matched/ unmatched) (SMD of 0.28 (95%CI 0.05, 0.51) compared to a SMD of 0.24 (95% CI -0.04, 0.53)).

Further analysis of two RCT datasets evaluated whether the use of two matched primary outcomes, as nominated by the trial authors, combined as a “co-primary” composite, might change the results and conclusions of these RCTs. These analyses demonstrated greater SMDs with the co-primary composites than the targeted composites in both Miyamoto et al. (136) and Moffett et al. (137). Comparing the RCTs with matched outcomes demonstrated that the composite outcomes were less likely to generate greater SMDs than the original matched outcome, in comparison to the unmatched outcomes where the matched composites in both RCTs generated greater SMDs that were more likely to be statistically significant in favour of the exercise arm.

5.4.1 How These Findings Compare to Other Studies

An increasing number of RCTs are now using more than one primary outcome (39% of included trials in the systematic review, chapter 3 section 3.3.5.2), and thus a composite combining the key primary outcomes may improve the ability of an RCT to detect differences in the effects of exercise versus control or comparison interventions over time (i.e. improve the responsiveness of the

outcome) (174,179). In the analyses of the trials with matched primary outcomes (136,137) the co-primary composite outcome generated a greater SMD than the matched composite, although the matched primary outcome showed greater SMDs than the composite in one trial at the primary time-point (136). This result suggests that including fewer matched key components in the composite outcome may improve responsiveness, perhaps because including more components leads to a dilution of effect (as some will change less than others from the exercise intervention). This improvement in responsiveness was noted by Parkes et al. (166) in rheumatoid arthritis patient data when comparing various composite outcomes with a single outcome, however, their analysis did not find improved statistical significance with composite outcomes in comparison to a single outcome (pain).

Other explanations for why the co-primary composite produced greater SMDs than the composite outcome may be that if an intervention does not equally address the different targets of the intervention, and the weighting of targets is not similarly reflected in the composite outcome, the composite may produce a result that falls short of a primary single matched outcome. In complex interventions, where multiple treatment targets are considered, weighting may be more applicable depending on how the intervention is delivered (175). This would require a weighting of the treatment targets of the intervention (e.g. 60% targeting pain, 20% targeting fear-avoidance beliefs, 20% targeting physical function), and then weighting the matched composite accordingly. The current limited understanding of the treatment targets and mechanisms of action of exercise in

NSLBP would make it difficult to justify such weightings. Further, this weighting and the complex calculations that accompany it would likely preclude such outcomes from being used in clinical practice.

It is important that in the creation of a composite outcome, components are reasonably combined, consistently defined and comprehensively reported (180). In this analysis, composite outcomes were generated based on the specified treatment targets articulated in the included example trials. Ross (181) recommends that composite outcomes should be related to the treatment targets of the intervention, have clinical meaning and be biologically plausible. In this study, all four RCTs had clearly stated their exercise treatment targets, and thus composites were 'RCT specific'. However, the future *a priori* identification of composite outcomes should in principle be influenced by a clear mechanism of change theory so that clinically relevant and meaningful outcomes are selected (52) that are directly related to the treatment targets of the intervention. The composites used in these exploratory analyses varied from four components (127) to eight components (137) (mean n=5). Each constructed composite was composed of different standardised outcome measures as identified by the RCT papers' authors. This prohibited comparisons of composite outcomes across datasets and would be a key barrier to use in RCTs, as the ability to pool data and compare results with other trials is important. The observed between-arm estimates of the effect of exercise versus control using the composite outcome may also be influenced by the responsiveness of the outcome measures that comprise it. In most of the analyses performed, similar outcome measures were

used to comprise the composite (e.g. RMDQ for physical function, TSK for fear-avoidance beliefs etc.). However, in some outcome domains such as physical function, although the RMDQ was most frequently used, the Patient-Specific Functional Scale (PSFS) was also used. Although both measures are known to be responsive, Hall et al. (182) report that the PSFS may be more responsive in patients with low activity limitation, and this may influence the responsiveness across different populations. Two different pain outcome measures were also used in the four included RCTs (the Brief Pain Inventory (BPI) and the numeric rating scale (NRS)). The BPI is known to have structural validity and internal consistency, but the NRS is preferred despite more inconsistent results on responsiveness (183). Although the BPI and PSFS outcome measures were used less frequently than the RMDQ and NRS, they may have increased coverage and therefore be preferable in a composite where a balance between responsiveness and coverage is desired. Although recommended core outcome measures for use in RCTs in persistent NSLBP have been agreed on (82), there is an acknowledgement that these measures are the best available, and not necessarily the best outcome measure for assessment (183).

Composite outcomes may be subject to misinterpretation if reporting is inadequate regarding the individual components. Further, the treatment effects may vary across different components of the composite outcome, or only a few components may change due to the exercise intervention which may result in the masking of a beneficial or harmful effect for one or more components of the composite (184). This again raises questions of what the RCT designers aimed

to target in their exercise intervention, what was actually targeted (see chapter 4, section 4.4.3), and what the treatment fidelity was of the treatment delivered. However, none of these RCTs reported a clear process evaluation, and thus it is difficult to understand the detail of the exercise interventions actually delivered.

There are many methods available to create a composite outcome. Creation of the composite outcomes in this study predominantly used the standardised averaging method described by Song et al. (175), but an alternative method is to use weight-adjusted averages. The field of cardiovascular research has used composite outcomes in varying compositions for many years, and recent reviews suggest 56% of cardiac surgery RCTs used a composite outcome as a primary outcome, with 14 different composites identified, highlighting the heterogeneity in their field (185). However, recent qualitative work has suggested that patients involved in cardiovascular research do not agree with the equal weighting of composite outcomes: patients stated they were more concerned about the effects of a permanent stroke-causing disability, than death (186). This suggests that future RCTs of cardiovascular research may need to incorporate a weight-adjusted composite to reflect patients' preferences (186). Weight-adjusted composites are more challenging to interpret and standardise, and would have been challenging to implement in the analyses used in this study as none of the included RCTs allocated percentages to the degree treatment targets were targeted with their interventions. If composites are to be used, they should be consistently defined, comprised of clinically relevant outcomes, and reported both as a whole and with individual components to ensure transparent

interpretation of results (187) which would be challenging for future research implementing a weighted composite. Further, this extends beyond the individual RCT wherein a composite was selected to reflect and match the specified treatment targets, but therefore would not allow comparisons between studies, and would limit the available data to be used in systematic reviews and meta-analyses. Core outcome domains have already been agreed to be used in all RCTs for persistent NSLBP, and the responder burden to participants of RCTs may become immense if RCTs include all three core outcome domains, matched outcomes and composite outcomes, unless the composite is comprised of many of these outcomes. This remains a key challenge in future RCT design and implementation. Future research should involve testing whether composite outcomes composed of outcomes matched to treatment targets might change the results and conclusions of trial datasets in a larger sample of datasets (n>4), such as of exercise interventions compared to no-exercise in an osteoarthritis population, or of manipulative interventions compared to usual care in a NSLBP population. Further research may analyse the precision and coverage of a composite outcome using principal component analysis, or alternative methods, to compare which method is best in creating a composite outcome with regard to effect size, coverage and standard error.

5.4.2 Strengths and Limitations of this Analysis

A strength of this exploratory secondary analysis was the acquisition of four RCT datasets, and the replication of the original analyses performed. The small number of RCT datasets analysed makes it difficult to draw firm conclusions;

however, it does provide initial evidence that a matched composite outcome in an RCT of exercise for persistent NSLBP can result in greater between-arm effect estimates and greater statistical significance than the use of a single outcome, although a co-primary composite outcome may generate greater SMDs than a single matched outcome. The analysis methods used replicated the primary analysis method used by the trial teams of the individual RCTs, and this ensured the results were comparable. These four RCTs were selected from a sample of twelve RCTs included in the systematic review, reported previously in this thesis, which may have been subject to publication bias. The limitation is that this was an exploratory secondary analyses of a small number of RCT datasets.

In all analyses performed in this secondary analysis the standard error of the composite variable was much smaller than that of the single primary outcome. A reason for this may be the improved precision related to the use of composites (192,193) and repeated measures of individual participants. Precision is likely enhanced (i.e. reduced standard error) through smoothing out measurement and random error as well as taking into account correlation between the outcomes comprising the composite outcome (188).

Reporting guidelines for composite outcomes recommend that the individual composite component results should be described with the results of the composite to ensure clarity (184,189). This allows the components of the composite outcome to be considered as secondary outcomes and allow for clear interpretation of the composite outcome results, which negates the need for

adjustment for multiplicity (184). A strength of this analysis is the reporting of individual components comprising the composite outcomes, however, due to the numerous outcomes included in the composite in some cases, and the word limits imposed by many journals, this level of reporting of composite outcomes is likely to be a challenge for RCT authors in future. It is recommended that the included components of a composite outcome should be related but not too highly correlated, or a single primary outcome domain would be more appropriate (184). This premise was not assessed in these analyses, and thus this may be a limiting factor when considering the results. The composite variable in these analyses was constructed with averaged standardised outcomes identified by the authors of each RCT publication.

5.4.3 Implications for Further Research

This study is exploratory in nature but nevertheless highlights that the results and conclusions of exercise RCTs may be changed by using a composite outcome that matches the targets of the exercise intervention. Further analysis of a larger number of RCTs of exercise interventions in persistent NSLBP is warranted. It would be useful to explore whether this is also the case for exercise interventions for different conditions, and for other complex interventions.

In RCTs of exercise interventions in persistent NSLBP, where the intervention may have multiple treatment targets due to its complex nature, a matched ‘targeted’ composite outcome may give the RCT team the best chance of detecting the benefits of exercise compared to control or comparison interventions. A composite outcome comprised of multiple matched outcomes to

the specified exercise treatment targets appears to generate a greater SMD in favour of exercise in comparison to a control arm. Well-matched co-primary outcomes may be preferable to a single matched outcome as they are more likely to cover a range of key treatment targets of exercise. However, identifying outcomes that are both responsive and have assess the right domains is difficult when the underlying treatment targets are poorly described and understood. The use of a matched composite outcome is likely to reduce the sample size requirements (166) for future RCTs which may have significant impacts on budget, resources and time taken to complete RCTs.

This study evidences the lack of consensus among trial teams about the key treatment targets of exercise. Even in the two RCTs investigating yoga as an exercise intervention (127,137), the authors identified many different targets (range from four to seven), of which only the reduction of pain was common to both. Thus, there is a clear need for a better understanding of the mechanisms of action of exercise (52) in order to be able to specify appropriate treatment targets that can be captured by well-matched and well-defined outcomes.

5.5 Conclusion

This exploratory analysis provides preliminary evidence that the use of a composite outcome, matching the multiple identified exercise treatment targets would have changed the results and conclusions of three of the four analysed RCT datasets. The use of composite matched outcomes in exercise interventions may provide a more responsive and meaningful outcome than a single matched

outcome. The next chapter develops this work further and aims to gain consensus on the treatment targets of exercise, and their agreed prioritisation for use in future RCTs of exercise interventions in persistent NSLBP.

6 Chapter 6: Gaining Consensus on the Treatment Targets of Exercise Interventions

Summary

This chapter describes two, sequential, nominal group workshops used to gain consensus on the treatment targets of exercise in persistent NSLBP.

The content of this chapter formed part of the peer-reviewed publication:

Wood L, Bishop A, Lewis M, Smeets RJEM, Bronfort G, Hayden JA, Foster NE. Treatment targets of exercise for persistent non-specific low back pain: a consensus study. Journal of Physiotherapy, accepted.

6.1 Introduction

There is a paucity of research that focuses on the treatment targets of exercise interventions in persistent non-specific low back pain (NSLBP). The systematic review that formed part of the earlier phase of this research programme (chapter 3) demonstrated that there are many and varied direct and indirect treatment targets of exercise described in published randomised controlled trials (RCTs) in the field of persistent NSLBP. The most frequently reported treatment targets were reducing pain, muscle strengthening, improving spinal stabilisation, flexibility and improving posture. The review observed that even in trials testing exercise programmes that follow a similar approach, such as the McKenzie method (125,137), the authors reported different treatment targets. This demonstrates clear uncertainty about the treatment targets of exercise, and the

underpinning mechanisms through which exercise benefits pain, function and quality of life (71).

Given this uncertainty, a logical next step is to undertake exploratory research that seeks to gain consensus about the treatment targets of exercise interventions with key stakeholders (190–192). In this study, key stakeholders were considered to be developers of exercise interventions that are tested in RCTs, people who use exercise to manage persistent NSLBP, and clinicians who prescribe exercise for this patient population. Nominal group workshop methodology was selected as the most appropriate approach with which to generate consensus, as it allows each participant an equal opportunity to present their ideas, discuss all ideas, and rank the generated topics in terms of importance. It is a time-efficient process as the panel of participants is kept focused and on track throughout the meeting (193). Furthermore, it allows confidentiality of ideas, as well as allowing the panel time to clarify thought processes and suggestions (193). This chapter describes the use of two sequential nominal group workshops to generate agreement on the most important treatment targets for exercise interventions in RCTs of persistent NSLBP.

6.1.1 Study Aim and Objective

Aim: The aim of this exploratory study was to identify and rank key treatment targets for exercise interventions in persistent NSLBP in order of importance.

Specific Objective:

To reach consensus on the important treatment targets of exercise interventions in persistent NSLBP by using nominal group workshops.

6.2 Methods

6.2.1 Study Design

Consensus was obtained by using the nominal group workshop technique in two separate, sequential workshops. Nominal group workshops were first described as a method through which committees make decisions by Van de Ven and Delbecq in 1972 (194), following research funded by NASA and the Institute for Research on Poverty for committees making decisions. The nominal group workshop methodology has previously been successfully used when incorporating multiple stakeholder groups together such as researchers, patients and clinicians, and as it ensures the opinions of all participants are identified and considered (195). The group is nominal in that it is highly controlled with discussion allowed at certain points through the process (196). Throughout the process, each participant has an equal opportunity to generate and present suggestions and cast their vote (and rank), preventing dominance of the group by some participants (195,197). Each nominal group workshop moves through five stages: introduction, idea generation, idea sharing, group discussion, voting and ranking (198), which were modified for each workshop. The final ranking stage generates consensus as a ranked aggregation of the participants' views rather than a communal viewpoint (199).

The nominal group workshop method was selected over the Delphi technique due to its participant interaction, relative speed of process and consensus generation, the high degree of task completion and the ability to gain valuable insight into discussion points around the ideas and question addressed (200). Reported disadvantages of nominal group workshops include the use of small numbers of participants (199).

One national workshop[11] was held, based at Keele University and a second international workshop was held at the International Forum for Back and Neck Pain Research in Primary Care in Quebec City, Canada. The initial plan for the PhD had only included one workshop (national workshop), but when the potential for a second workshop arose, the opportunity was taken to build on the results of the first workshop. Ethical approval was obtained from the University Research Ethics committee for both workshops: on the 17 August 2018 (reference ERP 2393) and an amendment was approved on the 15 April 2019 (reference MH-190019) (see Appendix 9.I: Additional Information Regarding Consensus Workshops). The methods for each workshop are described below. Both workshops used online systems to facilitate the workshop flow, and the national workshop used a mixture of both virtual and face-to-face mediums.

[11] The national workshop was initially the only workshop planned as a part of this PhD, and one international attendee responded to an invitation and took part. As the first workshop was attended by 93% UK attendees, this was named as National despite the inclusion of one international attendee in contrast to the second workshop where all attendees were from countries other than the UK.

6.2.2 Participants: Identification and Invitation

This study included researchers involved in the development of exercise interventions that are tested in RCTs, people who use exercise to manage persistent NSLBP, and clinicians who prescribe exercise for this patient population. The sample invited from each stakeholder group was a convenience sample. A limit of 25 participants per workshop was set to ensure manageable group discussion and input as part of the nominal group workshop process (197). For the first, national, workshop awareness raising for the nominal group workshop included: advertisements, telephone calls, targeted emails, social media, and contact with local patient groups for arthritis and back pain (details are presented in Table 6-1), in order to target researchers, people who had used exercise for their back pain, and clinicians (predominantly physiotherapists). A total of 15 participants provided informed consent for the first workshop. The second, international, workshop was held eight months later, as part of an international conference held in Quebec City, Canada. This workshop targeted researchers in the field of LBP who may or may not also be clinicians. The abstract for the workshop was submitted, peer-reviewed and accepted for the Forum conference programme. Several hundred attended the conference, and a total of 20 participants registered for the workshop. During the workshop, participants were able to freely join or leave the workshop at any stage.

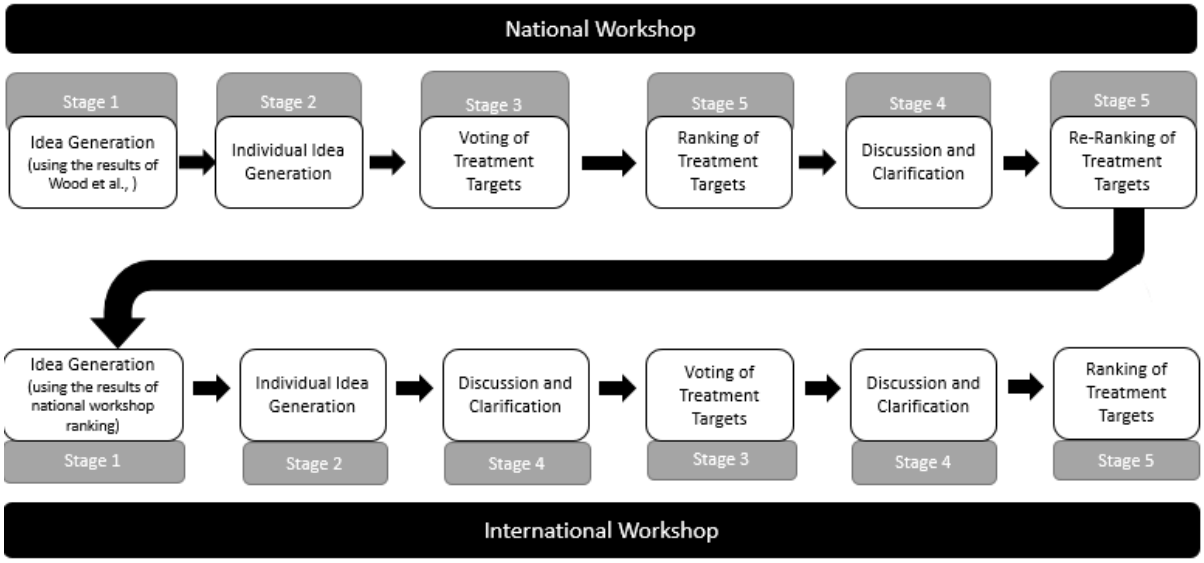
Table 6-1: Recruitment strategies used and their target population

Recruitment Strategy	**Participant Group Targeted**
Social media (twitter and Facebook)	People who have used exercise to treat persistent NSLBP Clinicians who have prescribed exercise for persistent NSLBP Researchers who have developed exercise interventions for persistent NSLBP
Email invitations to the BackCare and Arthritis Musculoskeletal Alliance Groups	People who have used exercise to treat persistent NSLBP
Advertisements were placed in the Chartered Society of Physiotherapy's Frontline magazine for the 2 months preceding the workshop	Clinicians who have prescribed exercise for persistent NSLBP
Advertisements with pull off strips were placed around Keele University campus	People who have used exercise to treat persistent NSLBP
Flyers were placed in waiting rooms of physiotherapists in the Staffordshire area	People who have used exercise to treat persistent NSLBP
Email invitations sent to authors of trials included in systematic review (chapter 3)	Researchers with experience of developing exercise interventions
Email invitations sent to individuals, centres and research institutes that were known to the researchers for previously developing exercise interventions tested in trials	Researchers with experience of developing exercise interventions

6.2.3 Nominal Group Workshop Process

Both nominal group workshops began with silent idea generation, as demonstrated in Figure 6-1. The national workshop was informed by the previous systematic literature review (chapter 3 in this thesis), and the results of the national workshop informed the international workshop. The process of achieving consensus through the initial voting, discussion and ranking stages followed published nominal group workshop guidance (197). The study process is demonstrated in Figure 6-1, and each of the stages is outlined thereafter.

Figure 6-1: Flow chart of consensus workshop stages (modified from Potter et al. (2004))



6.2.3.1 **National Workshop:**

6.2.3.1.1 ***Pre-workshop (Stage 1 and 2)***

Participants provided written informed consent (see Appendix 9.I: Additional Information Regarding Consensus Workshops) and basic demographic information using online data collection following their registration for the workshop. One week prior to the workshop, access to the online portal was provided through GroupMap (https://www.groupmap.com/) to view the list of pre-generated treatment targets informed by the systematic review (chapter 3, section 3.3.5.1). Participants were able to add suggestions to this list before the workshop, and participants' additions were private to the rest of the group until the PhD candidate enabled viewing permission during the workshop. Idea generation items that were similar were merged by the research team to avoid

duplication, and a total list of all potential treatment targets generated was created at the end of this process.

6.2.3.1.2 *Within Workshop (Stages 3-5)*

Participants were able to participate virtually through the use of Skype (https://www.skype.com/en/)(Skype Communications) or in-person face-to-face. All participants brought a mobile device to engage with the online platform during the workshop. The results of the idea generation stage were visually displayed to all participants (using the online platform), so each participant had the opportunity to add any omitted items to the list (as an extension of the original idea generation stage), and to facilitate discussion of the items displayed. Due to the range of participants (people with experience of using exercise to manage back pain, clinicians and researchers) and the number of additional items added in the second stage, discussion of the individual potential treatment targets was delayed until after the voting and ranking had reduced the number of targets. Participants then privately voted (yes/no) for their ten most important treatment targets (stage 3) using the online platform. The potential treatment targets that obtained at least one vote were carried through to the next stage and were presented to the group via the online platform. No thresholds for consensus were used within this workshop as it was the first stage in an exploratory study, so any target that received a ‘yes’ vote was carried through to the next stage. Items were then individually ranked in order of importance through an online ranking process (stage 5), where participants allocated ten points to their most important treatment target, descending to one point for the least important. The results of

this ranking process were discussed within the workshop (stage 4): treatment targets were discarded, grouped and amended as required through discussion and verbal agreement from the workshop participants. A further ranking stage was undertaken online two weeks after the workshop due to the extensive discussion and clarification stage.

6.2.3.2 **International Workshop:**

6.2.3.2.1 ***Pre-workshop (Stage 1 and 2)***

Two weeks before the workshop, registered attendees provided informed consent (see Appendix 9.I: Additional Information Regarding Consensus Workshops) and brief demographic information. They were also supplied with a list of the final ranked treatment targets from the national workshop and asked to add any other potential treatment targets that were absent using an online form developed with Google sheets.

6.2.3.2.2 ***Within Workshop (Stages 3-5)***

The results of the idea generation were presented to the workshop participants visually, allowing the opportunity for potential treatment targets of similar constructs to be clarified and grouped, discarded or amended as agreed by the group. Voting (stage 3) occurred individually on a document developed on Google Sheets (see Appendix 9.I: Additional Information Regarding Consensus Workshops), and participants were able to vote for any number of items they felt to be potential treatment targets (yes/no). In the voting stage, a pre-specified consensus threshold of 75% was set for item removal, where 75% or more of

participants agreed that the item was not a treatment target of exercise (201). These treatment targets were excluded from future voting and ranking rounds. Remaining targets were carried through to the next stage of further discussion and clarification (stage 4), followed by participants individually ranking the treatment targets in order of importance (stage 5). Ranking occurred in order of importance from one to ten, with results amended to reflect the same scaling as in the national workshop (i.e. the highest score was given to the one they felt was the most important). This ranking order was felt to be easier to implement at the international conference. The results were then presented to the participants of the workshop, who agreed with the order of importance of treatment targets, and declined the opportunity to re-rank treatment targets. These were then grouped into functional (relating to physical function), psychosocial (including psychological and social), behavioural (for example health service use) and impairment-based (such as strength or flexibility) targets (49) by the PhD candidate and supervisory team based on the nature of the treatment target.

6.2.4 Logic Model of Results

A logic model has previously been developed earlier in this thesis (chapter 3, section 3.2.7) to display the results of the systematic review, wherein the visual representation allows multiple relationships to be visualised, and the components of complex interventions to be more clearly described (114). The previously developed logic model reads from left to right and includes the setting of exercise delivery, the exercise deliverer, the treatment targets, the exercise specific components and the outcomes used in RCTs of exercise in persistent NSLBP.

The results of the consensus workshops summarised in this chapter of the thesis were used to update the treatment targets component of the logic model.

6.3 Results

6.3.1 Participants

A total of 15 participants took part in the UK national workshop (7 in-person face-to-face, 8 virtually using the GroupMap platform) the range of participants was 12 to 15 due to technical difficulties encountered by three participants which led to connection problems for part of the workshop. These three participants in the national workshop were unable to participate either due to connection issues or diary clashes. Twenty-four participants took part in the international workshop (with a range of 20 to 24 at different stages of the workshop). The range in number of participants in the international workshop resulted from participants moving in and out of the workshop due to the nature of a conference environment. Please see Table 6-2 for the baseline demographic information regarding participants. Across the two workshops, participants were from ten different countries (UK (n=15), Canada (n=6), Brazil (n=4), Finland (n=3), Norway (n=3), Australia (n=2), USA, Denmark, Sweden, Netherlands (all n=1)). 90% of participants had experienced back pain at some point in their lives.

Table 6-2: Demographic information of workshop participants

	National (%)	**International** (%)
Total Participants (n)	15 (range 12-15)	21 (range 20-24)*
Gender (Female n, %)	13 (87)	13 (62)
Participation Type:		
In person, face-to-face (n (%))	7 (47)	21 (20-24)
Virtual, via the GroupMap platform (n(%))	8 (53)	0
Profession:		
People who use exercise (n (%))	2 (13)	0
Researchers (n (%))	5 (33) **	21 (100)
Clinical Researchers (n (%))	2 (13)	
Clinician (n (%))	6 (40)	19 (90)
Type:		
Physiotherapist (n (%))	8 (100)	12 (57)
Chiropractor (n (%))	0	2 (10)
Medical Doctor (n (%))	0	4 (19)
Other Clinical (n (%))	0	1 (5)
Number of exercise interventions developed	(Of 6)	
<1	3	10 (48)
2-5	3	9 (43)
>5	0	2 (10)
People with lived experience of LBP	14 (93)	18 (86)
Different countries	2	10

*Range is represented due to the different number of participants across the stages. The percentage shown is of the total participants per workshop.
**All seven of the researchers involved in the national workshop had a clinical background or worked as clinical researchers.

6.3.1.1 **National Workshop Participants**

A total of six clinicians (40%), two people with a lived experience of LBP (13%), and seven researchers (47%) took part. Of these, two clinicians and one researcher were unable to join the workshop due to connection issues or last-minute diary clashes but they proceeded to complete the final ranking stage. Two people (who were neither researchers nor clinicians) who used exercise for their back pain took part, they had an average of 7 years (range 3.5 to 10+ years) of back pain experience to date, and both described their course of persistent NSLBP as ‘recurrent episodic’. Both of these people participated face-to-face in person in the workshop.

Six clinicians took part from a variety of locations across the UK. The average time they had been qualified was 18 (range 4 to 40+) years, with an average of 17 (range 4 to 40+) years involved in the treatment of patients with LBP. Four of the clinicians had not completed any formal post-graduate (PG) training course, two had MSc degrees, three had PG diplomas. All clinicians apart from one had experienced back pain at some point in their lives.

Five researchers and two clinician-researchers participated in the workshop from a variety of locations in the UK (n=14) and Canada (n=1). As seen in the methods (section 6.2.2, Table 6-1), researchers were invited from across the world, however only one international participant accepted. The majority had a physiotherapy background (five), with one chiropractor, and one traditional Chinese acupuncturist. The average amount of time involved in research was 12 (range 2-20+) years, and most had been involved with the development of two exercise interventions that had been tested within RCTs (range 0[12] to 4). Of the seven researchers involved, all had experienced back pain at some point in their lives.

6.3.1.2 **International Workshop Participants**

A total of 21 participants signed consent forms and completed the pre-workshop idea generation. During the session, three further participants participated in the

[12] Although researchers known to have developed exercise interventions were invited to participate, through invitations to research centres involved in exercise trials, and the systematic review authors, not all researcher participants had directly developed an exercise intervention.

voting stage, but only 20 respondents completed the final ranking stage due to movement in and out of the workshop during the workshop session. Almost all participants were researchers with a clinical background, who may or may not be clinicians as well as researchers at present (90.5%). The participants were from ten different countries, with the majority from Canada (22.7%) and Brazil (18.2%).

6.3.2 National Workshop Results

6.3.2.1 **Idea Generation (Stage 1-2)**

Fifteen participants completed the pre-meeting idea generation for the workshop. A further 25 different treatment targets were added to the original list of 30 targets, creating a total list of 55 treatment targets (please see Appendix 9.I: Additional Information Regarding Consensus Workshops for the full list).

6.3.2.2 **Voting of Treatment Targets (Stages 3)**

Discussion and clarification of the potential treatment targets were delayed until after the ranking stage (stage 5). Twelve participants voted for their ten most important potential treatment targets. At least half of the participants voted for ‘reducing pain’ (66.7%), ‘reducing fear of movement’ (50%) and ‘increasing functional capacity’ (50%). In total, 43 (of 55, 78%) potential targets received ’yes’ votes and were carried through to the next stage, as seen in Table 6-3. Items that received no votes and were discarded at this stage included: ‘reduce stress’, ‘improve proprioception’, ‘improve recovery’, ‘improve balance’, ‘reduce muscle tension’, ‘aid relaxation’, ‘restore neural mobility’, ‘reduce catastrophising’,

‘optimise neural function’, ‘improve posture’, ‘improve spinal stability’, and ‘improved control’.

Table 6-3: Treatment targets, number and percentage of votes (yes) after voting stage

Treatment Target (n=43)	**Number of votes**	**Percentage Vote**
Reduce pain	8	66.67%
Reduce fear of movement	6	50%
Improve functional capacity	6	50%
Increase function	5	41.67%
Muscle strengthening	5	41.67%
Increase physical activity and or exercise capacity	4	33.33%
Decrease barriers to movement	4	33.33%
Increase range of movement - spinal and other joints	4	33.33%
Enhance self-management skills	4	33.33%
Improve general health and well-being	4	33.33%
Reduce back pain	4	33.33%
Reduce other health services use (other treatments/ medication/ testing)	3	25%
Limit time to return to work (full, partial duties)	3	25%
Reduce anxiety and depression	3	25%
Improve mobility of the spine	3	25%
Prevent recurrence	3	25%
Improve fitness	3	25%
Improve strength	3	25%
Improve psychosocial factors	3	25%
Improve coping ability	3	25%
Improve self-efficacy	3	25%
Improve physical capability	3	25%
Decrease threat	2	16.67%
A tool to teach pacing and graduated increase in activity/ exercise	2	16.67%
Improve physical activity levels	2	16.67%
Reduce kinesiophobia	2	16.67%
Reduce absence from work	2	16.67%
Encourage normal movement	2	16.67%
Improve mobility	2	16.67%
Reduce disability	2	16.67%
Improve self-confidence	2	16.67%
Muscle flexibility	2	16.67%
Ensure mobility into the future	1	8.33%
Increase exercise compliance	1	8.33%
Improve work capacity	1	8.33%
Stretch the connective tissue	1	8.33%
Reduce the need for surgery	1	8.33%
Improve motor control	1	8.33%
Reduce dependence on health service	1	8.33%
Improve mental positivity	1	8.33%
Improve body awareness	1	8.33%
Reduce deconditioning	1	8.33%
Increase trunk muscle endurance	1	8.33%

Maximum number of votes 12.

6.3.2.3 **Ranking and Discussion of Treatment Targets (Stage 4 and 5)**

Participants then ranked their most important treatment targets in order of descending priority using the online platform. As the grouping stage had not been performed adequately in stage 4, significant overlap and replication remained across the list of treatment targets (e.g. most important was ‘reduce pain’, the second most important was ‘reduce back pain’). Three constructs were removed: ‘increase exercise compliance’, ‘improve body awareness’ and ‘improve psychosocial factors’. Discussion concluded that the use of ‘compliance of exercises’ to achieve a treatment target was important, but was not the goal of exercise in itself. ‘Improve psychosocial factors’ and ‘improve body awareness’ was felt to be too broad and not specific enough regarding the targets of exercise. Participants highlighted that they felt it important to have an equal number of physical and psychological targets in the top ten, and that these should include targets encompassing social function including work, and reducing reliance on other interventions. Therefore, further grouping of this list of treatment targets was undertaken. A reduced list of 18 unique treatment targets was developed from the original 43, (see Appendix 9.I: Additional Information Regarding Consensus Workshops for more detail). Examples of merged targets include the following: ‘reduce fear of movement’ was grouped with ‘decrease barriers to movement’, ‘decrease threat’, and ‘reduce kinesiophobia’ so that all of these targets were grouped together and called ‘reduce fear of movement’. Re-ranking of the 18 potential treatment targets was then performed online two weeks after the workshop with all participants who initially consented to participation (n=15), with each participant identifying their top 10 most important targets, by allocating

a score of 10 to the most important target and 9 to the next important and so on. The final ranking results are summarised in Table 6-4.

Table 6-4: Final ranked position of treatment targets from the national workshop

Rank Position	Treatment Target	Total Score	No. of participants that included this target in their top 10	% of participants with target in their top 3
1	Reduce pain	120	14	66.67
2	Increase function	109	15	60
3	Reduce fear of movement	100	15	46.67
4	Encourage normal movement	74	11	26.67
5	Improve mobility	60	9	26.67
6	Improve self-efficacy	56	10	20
7	Enhance self-management skills	54	10	13.33
8	Prevent recurrence	45	9	13.33
9	Improve general health and well-being	38	10	13.33
10	Improve strength	29	6	0
11	Increase exercise capacity	26	7	0.07
12	Increase physical activity	26	7	0.07
13	Improve work capacity	24	5	0
14	Reduce anxiety and depression	22	5	0
15	Improve motor control	17	4	0
16	A tool to teach pacing and graduated increase in exercise/ activity	16	3	0
17	Increase trunk muscle endurance	11	3	0
18	Reduce other health services use	2	1	0

The total score could range from 0 to 150. The maximal number of top ten rankings was 15.

Table 6-5 demonstrates ranking according to participant group. All three groups prioritised ‘reduce pain’ as their most important treatment target. The people who used exercise and the researcher group ranked ‘increase function’ in second place, while the clinician group ranked ‘reduce fear of movement’ over ‘increase function’. Interestingly, the constructs of ‘reduce anxiety and depression’, ‘improve trunk muscle endurance’, ‘increase physical activity’ and ‘increase work capacity’ were not reflected in the overall top ten ranking list although they featured within grouped rankings.

Table 6-5: Final ranked position of treatment targets per participant group in the national workshop

Rank Position	People who used Exercise (n=2)	Clinicians (n=6) and Clinician/Researchers (n=2)	Researchers (n=5)
1	Reduce pain	Reduce pain	Reduce pain
2	Increase function	Reduce fear of movement	Increase function
3	Prevent recurrence	Encourage normal movement	Reduce fear of movement
4	Reduce fear of movement	Increase function	Improve self-efficacy
5	Improve general health and well-being	Improve mobility	Enhance self-management skills
6	*Reduce anxiety and depression*	Enhance self-management skills	Improve mobility
7	Improve mobility	Prevent recurrence	Encourage normal movement
8	Improve work capacity	Improve self-efficacy	*Increase physical activity*
9	Improve self-efficacy	Improve strength	*Increase work capacity*
10	Enhance self-management skills	Improve general health and wellbeing	*Improve trunk muscle endurance*

Constructs 4 and 5 in the patient group, 3 and 4 in the clinician group and in the researcher group constructs 8 and 9 were tied and therefore share the same place. Items in italics are not reflected in the overall top ten priorities but were prioritised within sub-groups. Shaded blocks were common to all three participant groups.

6.3.3 International Workshop Results

Idea Generation (Stage 1-2)

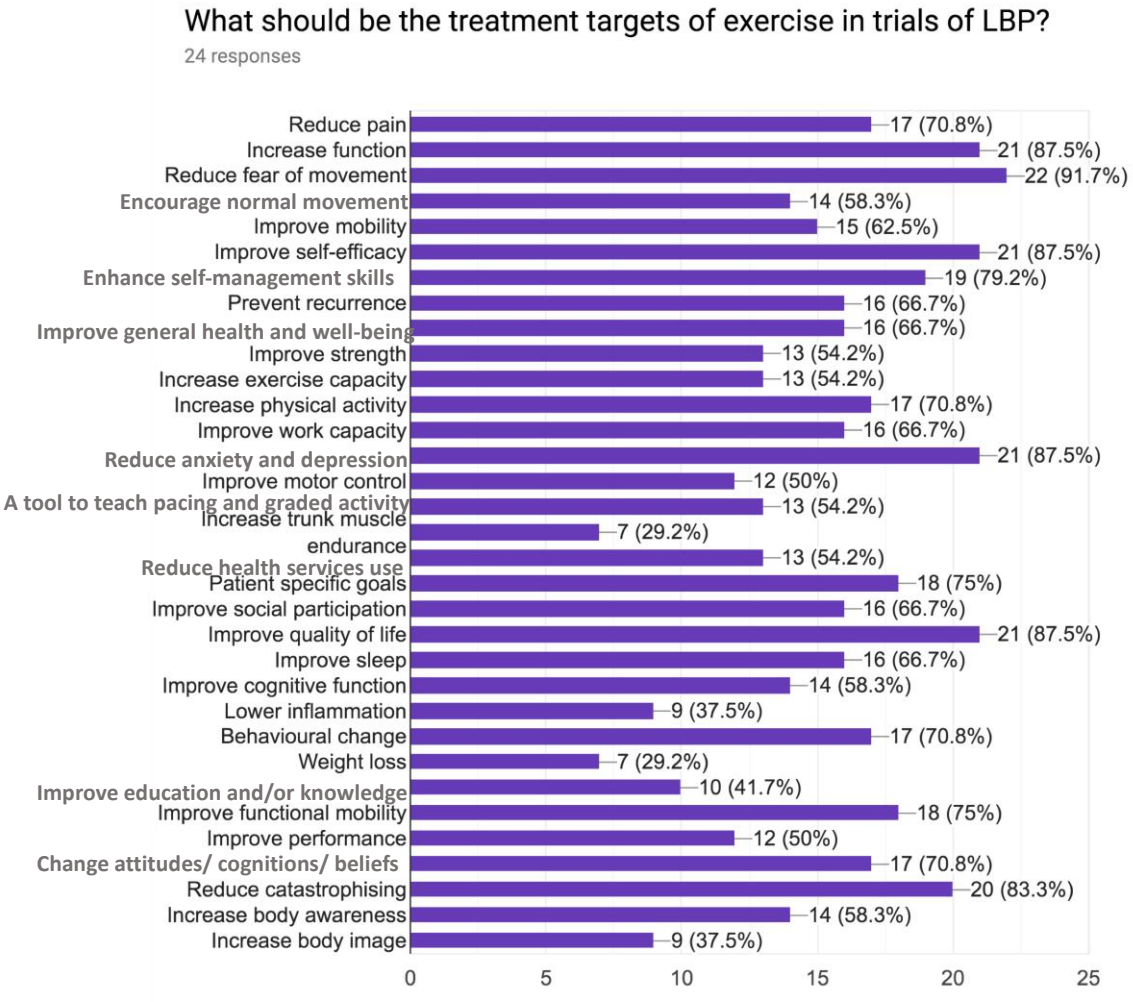
Twenty-one participants completed the pre-meeting idea generation for this workshop and added a further 15 targets to the original list of 18 treatment targets (see Appendix 9.l: Additional Information Regarding Consensus Workshops). This provided a total of 33 potential treatment targets for discussion, voting and ranking at the international nominal group workshop.

Voting and Discussion of Treatment Targets (Stages 3-4)

A total of 24 participants voted (yes or no) for treatment targets from the list of 33 potential treatment targets, within the international nominal group workshop. The

potential treatment targets that received the most 'yes' votes were 'reducing fear of movement' (91.7%), 'increasing function' (87.5%), 'improving self-efficacy' (87.5%), 'reducing anxiety and depression' (87.5%) and 'improving quality of life' (87.5%) as seen in Figure 6-2. No potential treatment targets were removed at this stage. Potential treatment targets of similar constructs were identified, discussed and grouped if felt to be appropriate by the participants, and discarded or amended as agreed by consensus of the group. Seven potential treatment targets were therefore removed at this stage.

Figure 6-2: Summary of treatment target votes in the international workshop (N and %)



Discussion centred on trying to group similar targets and clarify those that some participants felt were unclear. For example, some participants felt that physical exercise was a vehicle to elicit changes within patients and that communication was more important, with one participant saying “it is less about the physical training and more to do with the relationship…the non-specific effects…it is more about changing the patient’s ways of thinking…”. Some participants clarified the additional treatment targets they had suggested, an example was the treatment target of “performance” which a participant explained related to athletes who may have different targets from other (non-athlete) patient populations, but after discussion this was felt by participants to better fit with the target of ‘patient-specific goals’. Some raised their concerns that targets may be misconstrued in languages other than English. “Improved body awareness” was discussed as some felt it might be beneficial to change the word to ‘normalise body awareness’ instead of ‘increase’ body awareness, whilst others felt the key issue of importance was more ‘mindful movement’. ‘Work capacity’ was suggested by some participants to fit best within the domain of “function”, but other participants felt that ‘work capacity’ should be kept separate. The meaning of some targets, such as “perceived disability” and how it might differ to the meaning of the term “disability” were discussed and resulted in removal of this target, given participants felt that if kept in the list, many other treatment targets could be reworded to add the prefix of ‘perceived’. Instead, this was felt to be adequately reflected in the target “cognitions, attitudes and beliefs”. In total, six potential targets were either removed completely or collapsed within other treatment targets as part of the same target.

The temporality of treatment targets was discussed with the possibility of having proximal and distant treatment targets, for example, mediators of exercise interventions and outcomes from exercise, but this was felt to be difficult as there was not enough evidence to know which treatment targets lie on the causal pathway (i.e. which are mediators of the outcomes of exercise). For example, 'fear-avoidance' might mediate the pathway of persistent NSLBP to 'improved physical function' (202), so in this instance, 'fear-avoidance' may be the proximal treatment target, whilst 'physical function' might be the distant treatment target. Others suggested that some targets may be bidirectional, but there is not enough evidence to suggest which constructs were required to precede others, such as 'increasing physical function' to 'increase work capacity', or 'increasing work capacity' to 'increase physical function'. It was agreed that potential treatment effect modifiers should not be considered as treatment targets as these would be less likely to be open to alteration, in contrast to mediators which lie on the causal pathway. No treatment targets were removed or altered as a result of these discussion points. From the above stage, a total of 27 treatment targets were taken forward for individual ranking in order of importance using a pre-generated Google Sheets document.

Ranking of Treatment Targets (Stage 5):

In total, 25 of the 27 treatment targets received were ranked in order of importance[13]. A total of twenty participants ranked their top ten treatment targets by allocating the score of 1 to their most important treatment target and 2 to the

[13] 2 treatment targets received no ranking votes

next most important target and so on, the summary is displayed in Table 6-6. A shuffle function was used on the google sheets so that each participant viewed the list of 27 treatment targets in a different order. The top six most important treatment targets were ‘increasing functional ability’, ‘improving quality of life’, ‘reducing pain’, ‘patient-specific targets’, ‘reducing fear of movement’ and ‘increasing physical activity’. More psychological constructs (n=4) were prioritised than functional constructs (n=3) or impairment constructs (n=1) in the top ten. Participants declined the opportunity to perform a further ranking.

Table 6-6: Final list of prioritised treatment targets from the international workshop

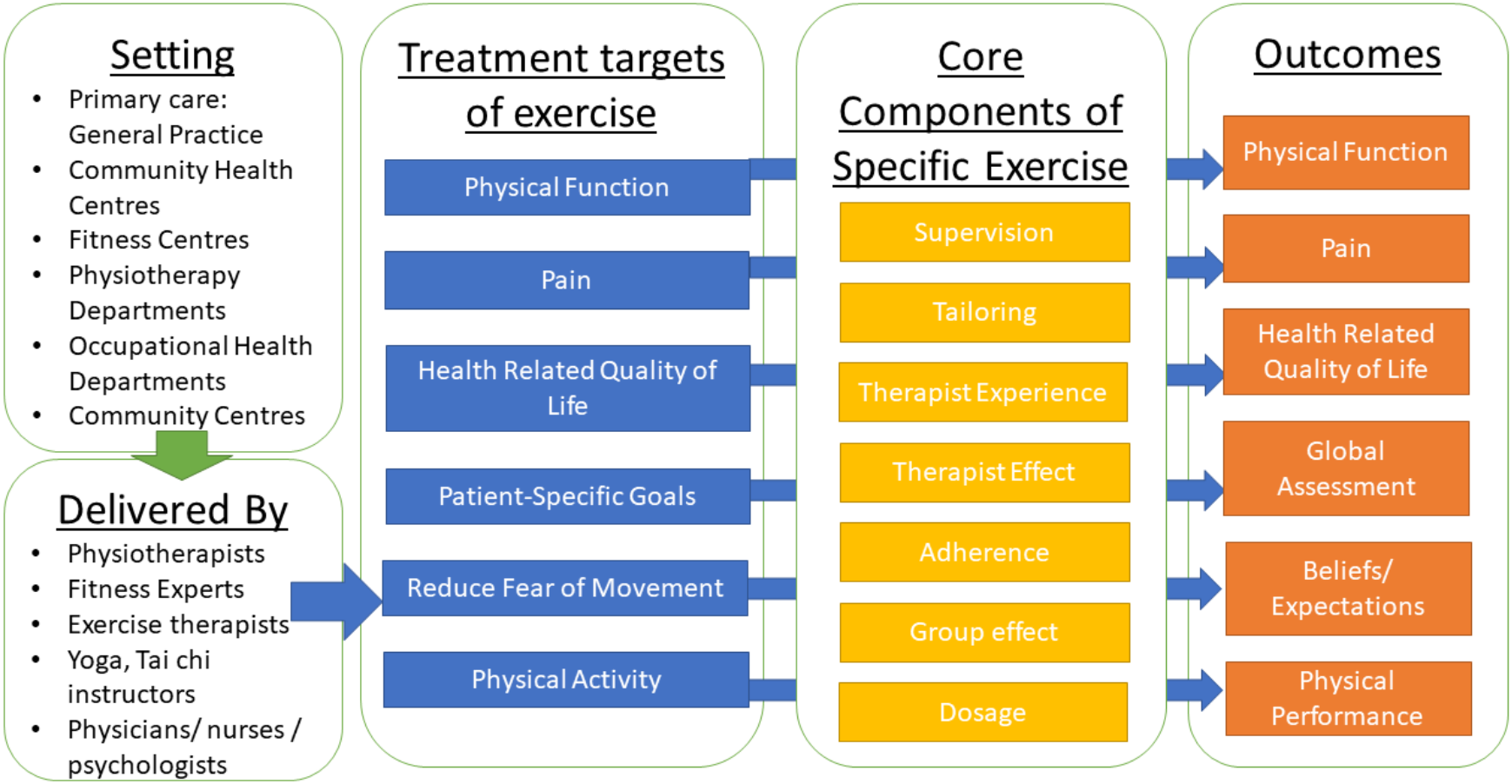
International Workshop (n=20 participants)				
Rank	**Treatment Target**	**Total Score**	**No. of participants that included this target in their top 10**	**% of participants with target in their top 3**
1	Increase functional ability	110	14	50
2	Improve quality of life	109	16	40
3	Reduced pain	91	14	40
4	Patient-specific targets/ goals	82	13	30
5	Reduce fear of movement	82	15	20
6	Increase physical activity	81	12	5
7	Improve self-efficacy	71	13	20
8	Enhance self-management skills	66	12	5
9	Improve work capacity	53	12	10
10	Improve attitudes/ cognitions/ beliefs	48	9	5
11	Reduce catastrophising	44	9	5
12	Prevent recurrence	42	9	10
13	Improve social participation	35	5	10
14	Reduce anxiety and depression	25	5	0
15	Reduce other health services use	24	7	0
16	Improve sleep	23	5	10
17	Allow behavioural change	16	5	0
18	Increase body awareness	15	3	0
19	Improve strength	14	3	0
20	Increase exercise capacity	13	4	5
21	Improve motor control	13	3	5
22	Increase body image	8	1	5
23	Improve spinal mobility	4	2	0
24	Weight loss/ gain	3	1	0
25	Increase trunk muscle endurance	2	1	0
26	Improve cognitive function	0	0	0
27	Lower inflammation	0	0	0

Treatment targets are broadly grouped into blue (functional), green (behavioural), orange (impairment-based), and pink (psychosocial) constructs. Total score ranged from 0 to 200 (0=worst attainable score, 200= best attainable score); number of top ten rankings ranged from 0-20 (0= least, 20=highest).

6.3.4 Updated Logic Model

The results of the final nominal group workshop were used to update the initial logic model created by the results of the systematic review (chapter 3, section 3.3.9). This may provide an example model for future trials to use when considering the development and specification of their exercise interventions. As seen in Figure 6-3, it includes the top six most highly ranked treatment targets, which largely map onto the six primary outcome domains used most frequently by RCTs of exercise interventions in persistent NSLBP.

Figure 6-3: Updated logic model including prioritised treatment targets of exercise



6.3.5 Summary of Results

Two sequential nominal group workshops were held with a total of 39 participants from ten countries. In total, 40 treatment targets of exercise for persistent NSLBP were generated across both workshops. Eighteen treatment targets were brought forward from the first, national, workshop, and by the end of the second workshop, 27 treatment targets were ranked as most important. The top six most important treatment targets were ‘increasing functional ability’, ‘improving quality of life’, ‘reducing pain’, ‘patient-specific targets’, ‘reducing fear of movement’ and ‘increasing physical activity’. An updated logic model incorporated the six most highly ranked treatment targets to present a visual, example, guide for future trials to consider when specifying their exercise intervention, including the treatment targets and outcomes.

6.4 Discussion

This study is the first to attempt to develop a consensus about the treatment targets of exercise for NSLBP with key stakeholders. The two sequential nominal group workshops generated consensus from a group of people with lived experiences of NSLBP, clinicians and researchers involved in prescribing or developing exercise interventions in clinical practice and in RCTs of persistent NSLBP. Many and varied targets were identified, indicating the challenge to reach consensus on a small number of treatment targets. From an initial list of 30 treatment targets informed by a systematic review (chapter 3), participants generated 40 treatment targets, and ranked a final list of 27 treatment targets. With 39 participants from ten countries, this consensus study found that

'increasing functional ability', 'improving quality of life', 'reducing pain', 'patient-specific goals', 'reducing fear of movement' and 'increasing physical activity' were the most important treatment targets of exercise interventions for patients with persistent NSLBP. Psychosocial constructs (40%) and functional constructs (30%) were prioritised in the top ten treatment targets.

6.4.1 Prioritisation of Treatment Targets

In terms of comparison between the results of this study and those from previous research, the top three prioritised treatment targets ('improving function', 'quality of life' and 'pain reduction') are similar to the treatment targets of exercise for persistent NSLBP suggested by Rainville et al. (49) (addressing functional impairments, reducing pain and back pain-related disability). Rainville et al. (49) suggested these treatment targets after performing a systematic review, whereas this study was the first to include stakeholders as well as a systematic review. However, in the systematic review earlier in this thesis (chapter 3), the most frequently reported treatment targets identified in RCTs of exercise interventions for persistent NSLBP were 'reducing pain', 'muscle strengthening', 'improving spinal stabilisation', 'flexibility' and 'improving posture'. Of these, only 'pain reduction' is reflected as a treatment target of importance in this consensus study. 'Improving strength' was ranked at number 19 out of 25, 'improving motor control' (incorporating 'spinal stabilisation') was ranked at 21 out of 25, 'improving spinal mobility' (including 'improving flexibility') ranked at 23 out of 25, and 'improving body awareness' (grouped with 'improving posture') ranked at 18 out

of 25. This demonstrates a potential shift in contemporary understanding of how exercise may work to reduce NSLBP and improve function and quality of life. However, this may also be the result of the different approach to identifying the targets (one using a systematic review of published trials, compared to this approach of consensus from stakeholder groups). There may also be differences in the prioritised treatment targets due to the role of the different stakeholder groups. However, in the first workshop, both 'reducing pain' and 'increasing function' were prioritised by all three stakeholder groups in first and second position of importance, suggesting that these have similar importance irrespective of stakeholder group, where this group included a large amount of research active clinicians.

'Reducing pain', 'improving function' and 'quality of life' are well-established as important core outcome domains and are the most commonly reported primary outcomes in RCTs of persistent NSLBP (68,70). When comparing these results, it can be seen that these are the same variables – providing evidence that the most important treatment targets of exercise may be those prioritised by the core outcome domains. However, these are the core outcome domains for *any* trial of *any* intervention, not necessarily exercise-specific. Perhaps the dominance of the research active clinicians in this consensus exercise who were more likely to be aware of the core domains recommendation were influenced by this knowledge in their decision-making. Further, the discussion element of the nominal group workshop may have meant that those who were not researchers/ clinicians may have been influenced by the discussion element. This suggests that the uptake

and acceptance of the core outcome domains by both clinicians and researchers has been widespread, and remain a high priority for people with NSLBP, but their role as treatment targets (or mediators) of exercise interventions remains unclear.

Although psychological treatment targets were not prioritised in the top three targets, a higher proportion of psychological treatment targets was ranked in the final workshop (40%) than in the first, national workshop (30%) or from the systematic review (10%) (see 3.3.5.1 in Chapter 3). The psychological constructs of 'reducing fear of movement', 'improving self-efficacy' and 'reducing anxiety and depression' received 'yes' votes from many participants in the second, international workshop. However, 'reducing anxiety and depression' was not ranked in the top ten priorities, despite the evidence for exercise in treating depressive symptoms (203). These psychological targets may feature more strongly in the international nominal group workshop due to the increased participation of clinical researchers, who may be more familiar with the literature on mediators of the outcomes from exercise in NSLBP. The psychological benefits of exercise are well-documented (36,145), so the prioritisation of psychological targets over other physical performance targets is perhaps not surprising. As the aetiology of persistent NSLBP is increasingly understood to comprise psychological, mechanical and social components (147), the emphasis placed on the psychological targets of exercise for persistent NSLBP seems to make sense. However, they were not prioritised as the most important targets for exercise, suggesting that these 'core outcomes' are still seen as the most

important to focus on. Both reducing fear of movement and self-efficacy have been shown to partially mediate the relationship between pain and disability in persistent NSLBP (204,205). Recent research on the mechanisms of exercise for persistent musculoskeletal pain suggests reducing fear of movement, improving self-efficacy, enhancing self-management strategies, and improving belief, cognitions and attitudes (all in the top ten treatment targets in the final workshop) may explain how exercises benefit patients (206). Perhaps surprisingly, although reducing pain catastrophising has been shown to mediate physical function with exercise in NSLBP (150), this was ranked in eleventh place. Increasingly, mediation analyses on these and other psychological constructs are being conducted as part of RCTs (45,148,152,207,208) to better establish whether these treatment targets lie on the causal pathway between the exercise treatment and the outcomes most important to patients (most commonly pain and disability) (209).

Including the three different stakeholder groups of developers, prescribers and users of exercise, in the first national nominal group workshop, was important to ensure all perspectives of exercise for persistent NSLBP were considered. All three of these groups prioritised similar treatment targets in their top four: ‘reducing pain’ was unanimously the most important treatment target, followed by ‘increasing function’ (second in people who used exercise and researcher groups), and ‘reducing fear of movement’ (second in clinician group, third in researcher group and fourth in people who used exercise group). These three targets were all prioritised within the second, international workshop within the

top five (‘pain reduction’ ranked third, ‘increase in function’ ranked first, and ‘reduction of fear of movement’ ranked fifth). People who used exercise placed importance in the first nominal group workshop on preventing the ‘recurrence of back pain’ and ‘reducing anxiety and depression’, which were not similarly ranked by other stakeholder groups. This suggests that there may still be a discord between the treatment targets of exercise prioritised by people who use exercise that were not captured within this final, international workshop. However, these two nominal group workshops consisted of participants from various stakeholder groups, which has been shown to improve the diversity of idea generation and richness of discussion (193).

In the final workshop ranking, no great consensus in the prioritised targets was achieved. Traditionally, many consensus workshops do not use consensus thresholds (197,198,210,211) as the overall ranking result is an amalgamation of the aggregate consensus. In contrast, some studies have used a threshold of 50% during the voting stage (meaning that each construct would require at least 50% of participants to vote yes), in order for a construct to be carried forward to the ranking round (212). If this approach had been used in the national workshop, very few treatment targets would have been brought forward to the ranking stage, as only three treatment targets had more than 50% ‘yes’ votes (‘reducing pain’ (66.7%), ‘increasing physical function’ (50%), and ‘reducing fear of movement’ (50%)). However, in the international workshop, many more treatment targets had more than 50% ‘yes’ votes, with 28 of 33 (85%) possible treatment targets gaining 50% or more, suggesting greater consensus amongst that group. This

difference may also reflect the difference in methods between the two workshops – in the first, national, workshop, participants were asked to vote for ten targets they felt were important, whereas in the second, international, workshop, participants could vote for as many targets as they felt were important. A consensus threshold of 75% (agreeing that the treatment target was not a priority in exercise for NSLBP) was used in the international workshop, however, no treatment targets met this threshold, and all were carried through to the ranking stage. This could perhaps have been a function of the sequential nature of the two workshops (given that several targets had already been removed, through consensus, in the first workshop). Reviewing the final ranked targets in the international workshop, only 'increasing function' was included in 50% of participants' top three constructs, and 'reducing fear of movement' and 'improving quality of life' were included in the most individual top ten priorities (ranked by 15 and 16 out of 20 respectively). This suggests that the overall consensus developed about the importance of the treatment targets was not very high. This may have happened given the general use of the term 'exercise intervention': participants may have found it challenging to reach agreement on the treatment targets for exercise in its broadest sense, as opposed to very specific exercise interventions, however this was not reflected in the discussion comments. Had a specific exercise type been selected as the focus, it may have been easier to reach agreement across the group. This would have then required the input of stakeholders who had developed, prescribed and used that specific type of exercise for NSLBP, an approach that may be useful for future research.

Other consensus studies have reviewed the individual votes per treatment target in the final ranking stage, by assigning median scores to the overall rank to suggest appropriateness (or importance in this instance) to assess whether consensus was present on the importance of the target (213,214). Applying this method in the current study may have altered the overall ranked positions. The method selected was the recommended method for use in nominal group workshop technique, and represents an overall consensus of the group (198) which has been shown to have validity (215).

6.4.2 Strengths and Limitations of this Study

This consensus study is the first to develop consensus about the treatment targets for exercise interventions in persistent NSLBP with a variety of key stakeholders involved in the use of, development, evaluation and prescription of exercise interventions. A further strength is the sequential approach taken from the systematic review into the two sequential workshops, which ensured refinement of the potential treatment targets. Consensus was generated with a large number of stakeholders (n=39), including people with lived experience of using exercise to manage persistent NSLBP, clinicians prescribing exercise for persistent NSLBP, and researchers who design and evaluate exercise interventions in RCTs from a number of countries. A limitation of the research conference used for the international workshop meant that the clinicians involved were largely research active clinicians, which may not therefore reflect the views of the many clinicians who manage patients with NSLBP who are not interested in nor active in research. This may also explain why the eventual consensus

reached on the top six targets includes the three recommended core outcome domains for research trials in the field of LBP, irrespective of the treatment being investigated.

A limitation of this consensus study was that it included only two people, who were not researchers nor clinicians, with lived experience of persistent NSLBP due to lack of response from invitations to patients and the public. This led to imbalance of the views represented by people who had a lived experience of persistent NSLBP in the first workshop (2 compared to 7 researchers and 6 clinicians). However, most of the researchers and clinicians who took part had personal experience of back pain. Although this participant group included clinicians and researchers from ten different countries, most participants were from the discipline of physiotherapy and from higher-income countries with well-established health-care systems (there were only four medical doctors included for example) and none from low and middle-income countries. During the international workshop, some participants left or arrived between each stage of the nominal group workshop, resulting in different numbers of participants completing the idea generation, voting and ranking stages. Due to the lack of identifiable data collected, it was not possible to trace which individuals completed each stage. Like all consensus studies, the results may have been different had different stakeholders been invited and participated.

Although participants in the online environment may feel less inhibited, it has been recorded as being 40% less efficient than the traditional method (216). In

this study, we had three participants drop out due to technology issues in the first workshop which may have had an impact on the overall voting and ranking stages as these participants were absent from the discussion and voting sections. However, due to the final ranking stage completed electronically after the workshop, all participants were able to complete this, which negates the impact the technological issues raised. The time to train and teach some of the participants in how to use GroupMap during the first workshop may have hindered the timekeeping, which resulted in an inability to group items during the workshop itself. Sending out pre-workshop training on the use of the platform may help to prevent problems such as this in the future (216).

6.4.3 Implications for Future Research and Clinical Practice

This consensus study has identified the most important treatment targets for exercise interventions in persistent NSLBP, with priorities including ‘improving function’ and ‘quality of life’, ‘reducing pain’, ‘targeting patient-specific goals’, ‘reducing fear of movement’ and ‘increasing physical activity’. Overall, 25 targets were ranked and prioritised, with consensus about the most important overall (top five targets), which may be useful for considering within future research data collection and mediation analysis within trials, helping to further understand how exercise works for patients with NSLBP. These agreed treatment targets may guide the design of RCTs of exercise interventions for persistent NSLBP by helping to target exercise interventions to achievable, measurable outcomes that match the aim(s) of the intervention. Future trial design may benefit from the use of intervention logic models to map out the role of treatment targets and select

the most appropriate outcome domains and measures for complex interventions, such as exercise, with multiple intervention targets (13,17). The identified treatment targets may, therefore, help in the identification of potential mediators of exercise that should be measured in future studies and used in pre-specified mediation analysis within RCTs.

Consensus work can be limited by the involvement of stakeholders in each workshop, although the aim of the consensus workshops is not to be generalizable. Given the change in priorities of the treatment targets identified in the two workshops, future workshops would benefit from greater involvement from people with a lived experience of persistent NSLBP. Further consensus work with different stakeholders comprising similar groups to those included in this study (i.e. researchers who develop interventions, clinicians who prescribe and people who use exercise for NSLBP) from different backgrounds such as low- to middle- income countries may add to the robustness of the results. Similarly, a repeat of this consensus study in a different format such as the Delphi method may provide different results to these presented here, as the discussion element and influence from other members in the group will be removed, which may have possibly influenced the final voting and ranking stages of this study. Further, a focussed consensus workshop on a specific exercise type (e.g. such as yoga or pilates) including various stakeholders to agree the specific treatment targets may be useful in designing trials to address these agreed, prioritised targets.

The logic model incorporating the six top ranked treatment targets which map onto the six primary outcome domains used by RCTs of exercise for persistent NSLBP has purposefully been developed to include multiple potential pathways, which can then be refined or amended for use in trials in the future. This is expected to vary, dependent on the exercise intervention to be tested, as well as the treatment targets identified by the trial team, and may in turn affect the specific exercise components as well as setting and deliverer considerations.

These targets may be addressed individually when identified at assessment or alongside multiple treatment targets when prescribing, designing, and/or evaluating exercise interventions for persistent NSLBP. By having a greater understanding of the comparable importance of these treatment targets, as viewed by a group of different stakeholders, clinicians may be encouraged to target their exercise intervention to ensure they select the most appropriate exercises to get the most benefit for their patients. However, there is still little known about the underlying mechanisms of exercise to affect pain and physical function, but as a clinician, there is a growing likelihood that targeting the treatment targets identified in these workshops may be of improved benefit to patients. As clinicians, by assessing patients for potential treatment targets such as strength deficits, fear-avoidance, self-efficacy, etc., they can identify priorities for their exercise treatment to target, and thereby select the most appropriate exercise. This can be performed within the framework of shared-decision making, to allow the selection of the most appropriate treatment targets for each patient.

Further, the advances of technology allowing objective outcome measurement in the forms of physical activity tracking, step counts, heart rate levels for example, provide many more cost-effective options for future triallists to consider when designing trials with matched outcomes. Given physical activity was ranked at sixth in importance, it is a treatment target that is of importance and future trials may need to be designed to capture this information through more innovative means.

6.5 Conclusion

This consensus study incorporating 39 key stakeholders from ten countries, involved in exercise use, prescription, delivery and evaluation, added 40 treatment targets of exercise for persistent NSLBP, ranked 27 of those as the most important, of which the six most important were: ‘increasing function’, ‘improving quality of life’, ‘reducing pain’, ‘targeting patient-specific goals’, ‘reducing fear of movement’ and ‘increasing physical activity’ as the most important treatment targets of exercise interventions in persistent NSLBP. Researchers testing the effectiveness of exercise, should be aware that exercise has multiple treatment targets, be clearer about which specific treatment targets their exercise intervention is focussed on, and consider how they might use this information to optimise patients’ outcomes from exercise. Clinicians prescribing exercise for NSLBP may use exercise to target the top ranked treatment targets. Future studies that test the identified treatment targets (such as by using a logic model, as demonstrated within this chapter) that are designed to find out if these are indeed mediators of patients’ outcomes from a specific exercise programme

– either as an individual treatment target or as part of multivariate mediation models – are required.

Further research will be useful to repeat this work, possibly with a different method such as Delphi that may limit the discussion opportunity, with a broader range of clinicians and exercise prescribers, and people who use exercise with NSLBP, as well as in settings other than high income countries, and that is focussed on more specific exercise approaches or types.

7 Chapter 7: Discussion, Summary and Conclusions

Summary

This chapter summarises and discusses the findings of this doctoral programme of research and makes recommendations for future research and clinical practice.

7.1 Introduction

This discussion draws together all the stages of this programme of research, summarised in the previous chapters of this thesis. The overall aim of the research was to explore whether better matching of the outcome domains used in RCTs in the field of persistent NSLBP to the treatment targets of the exercise interventions might change the results and conclusions of these trials. Having identified the gaps in knowledge to date (chapter 1), this thesis aimed to understand whether existing RCTs of exercise interventions in persistent NSLBP matched their primary outcome domains to the identified (if any) treatment targets of the exercise intervention by systematically reviewing the available literature (chapter 3). Within this review, RCTs were identified where secondary analyses could be undertaken to understand whether matching of primary outcome domains to exercise treatment targets might alter the observed between-treatment arm differences and, therefore, the conclusions of these RCTs (chapter 4). In chapter 5, composite outcomes composed of multiple matched outcome domains were compared to single primary outcome domains to evaluate whether

these altered observed between-treatment-arm differences and RCT conclusions. Finally, this thesis aimed to generate consensus about, and prioritise treatment targets of exercise interventions for persistent NSLBP through the use of nominal group workshops (chapter 6). The key findings of this thesis are summarised below and set in the context of the most recent relevant research, particularly more recent RCTs of exercise for persistent NSLBP. The chapter concludes with key implications and recommendations for both research and clinical practice.

7.2 Summary of Thesis Findings

7.2.1 Systematic Review

Existing guidance for the design of RCTs recommends using a primary outcome measure that is appropriate, given the key variable that the intervention is trying to change (83). In the field of NSLBP there is a paucity of literature that describes or defines treatment targets for exercise interventions. A systematic review was undertaken to identify what treatment targets and outcomes have been used by the authors of previous RCTs testing the effectiveness of exercise interventions for persistent NSLBP, where only trials with a minimum sample size of 60 participants per arm were included. Within this, RCTs were categorised as ‘matched’ if they had used a primary outcome which reflected the treatment targets identified by the RCT authors in their RCT publications. The results of these RCTs were then compared to the ‘unmatched’ group in which the primary outcome was not judged to be reflective of the treatment targets identified.

A total of 27 RCTs were included in the systematic review, and only 27% of RCTs were categorised as ‘matched’, suggesting that most exercise RCT teams have previously not selected their primary outcome based on the way it matches the treatment targets of their exercise intervention. ‘Matched’ RCTs were more likely to show a statistically significant between-treatment arm difference in favour of exercise in comparison to a non-exercise control (71% of the ‘matched’ trials compared to 15% of the ‘unmatched’ trials). When comparing the SMD of ‘matched’ to ‘unmatched’ RCTs, the ‘matched’ RCTs demonstrated a greater SMD (SMD 0.54 (95% CI 0.23 to 0.85) (p=0.0006)) than the ‘unmatched’ RCTs (SMD 0.22 (95% CI 0.01, 0.44), p=0.04), although this difference was not statistically significant. This finding was replicated in sensitivity analyses including ratio of means, weighted mean differences for pain and physical function; and sub-group analyses of comparator groups, recruitment strategies, risk of bias and specification of treatment targets. This exploratory review provided some initial support for the hypothesis that better matching of outcome domains to the treatment targets of exercise interventions in RCTs with patients who have persistent NSLBP may be more likely to lead to statistically and clinically significant results, in favour of exercise compared to non-exercise controls.

This systematic review also identified that most RCTs used similar primary outcomes, favouring physical function and pain domains (used in 56 and 52% of included RCTs respectively). A wide variety of secondary outcome measures was used, with an average count of five outcomes per RCT. Thirty-one different treatment targets were identified across the set of 27 RCTs of exercise in

persistent NSLBP, the majority of which were only cited by one RCT. Only 66.7% specified their treatment targets, of which 22.2% inferred treatment targets indirectly in the introduction or methods sections of their RCT publications. Despite this review using more stringent inclusion criteria of larger sample sizes, the findings are similar to a recent review of 403 exercise arms included in 265 RCTs, in the recent update of the Cochrane exercise trials for LBP. This review found that only 59% of RCTs specified treatment targets, of which 23% were inferred (67). The most frequently cited treatment targets of exercise were improving muscle strength, spinal stabilisation, reducing pain, improving flexibility and improving posture; again, similar to that found in the recent Cochrane review (67).

A logic model was constructed based on the results of the treatment targets and outcomes used in the set of included RCTs. However, due to the high level of heterogeneity in the treatment targets and outcomes used by RCTs for NSLBP and the combination of treatment targets and outcomes used within each trial, it was not possible to identify any clear relationships between and across RCTs in the outcomes and most frequent treatment targets used. All of these results, from the systematic review, highlight clear uncertainty regarding the treatment targets of exercise for NSLBP, with RCTs testing similar exercise interventions identifying different treatment targets and using different outcomes, and providing different rationales for their choices. Exercise is a good example of a complex intervention, and RCT teams might benefit from using logic models to map out and articulate more clearly their intervention rationale, intervention targets and

proposed mechanisms of action, intervention components, and relationship between the treatment targets of the intervention and outcomes.

This review highlighted the poor matching of the primary outcome to the identified treatment targets of the exercise intervention in previous RCTs for persistent NSLBP. RCTs that matched the primary outcome to the treatment targets of their exercise interventions were more likely to demonstrate a statistically significant between-treatment arm result in favour of exercise and to report larger SMDs than those that did not match their primary outcome. Therefore, matching this could be an important consideration for future RCT teams to consider as they develop the plans for their exercise intervention and outcome selection. RCTs are the best design to compare the effectiveness of complex interventions such as exercise, as they reduce the potentially biasing effect of confounders through the randomisation process (55). As the Medical Research Council (52) describe, “a good understanding of how the intervention effects change is required to ensure any weak causal links can be identified and strengthened”.

Given that the end date of the search within the systematic review described in chapter 2 was August 2019, a brief updated search was undertaken to inform this discussion on 1 June 2020. The findings showed that of 23 RCTs currently in progress, registered on USA and UK RCT registries and testing exercise interventions for persistent NSLBP, only ten (43%) specify their treatment target(s). Of these, only four (17%) use a primary outcome that reflects their treatment target(s) (please see Appendix 9.m:Summary of Protocols of 23 Current RCTs). This suggests the systematic review results in chapter 3

continue to reflect current RCT practice and therefore, the findings and implications continue to be relevant.

7.2.2 Data Analysis 1: Matching Outcomes to the Treatment Targets

This secondary data analysis sought to explore whether matching the primary outcome(s) to the identified treatment targets of exercise interventions changed the results and conclusions of existing RCTs of exercise for NSLBP. Five RCTs from those included in the systematic review were selected for inclusion in this secondary data analysis, of which, three RCT datasets were obtained (total of n=1033 participants). The other two RCTs were also analysed, but, using information from their published papers, comparing the SMD for the primary outcome(s) to the SMDs of the matched outcomes with data described within the two publications.

Of the five different analyses, three sets of results showed that the use of matched outcomes in comparison to unmatched outcomes demonstrated greater between-arm SMDs (and three of the four showed statistically significant results in favour of the treatment arm but not between the outcomes), in favour of the exercise intervention in comparison to a non-exercise control arm. Secondary analysis of only one of the five RCTs showed a greater SMD in favour of exercise, versus control, when using an unmatched primary outcome. Using the first mentioned treatment target from each included RCT to generate a summary forest plot demonstrated a greater SMD using the matched outcomes (0.30 (95%CI 0.04,0.56) (p=0.02)) compared to the unmatched primary outcomes

(SMD 0.19 (95% CI -0.03, 0.40) (p=0.09)), although the difference between these sub-groups was not statistically significant (SMD 0.11 (95%CI -0.34, 0.57)).

In complex interventions where there may be multiple treatment targets (14,15), identifying the primary treatment target and therefore the primary outcome can be challenging. Thus, it may be more appropriate to use a composite outcome consisting of all the outcomes that match all the specified treatment targets. As seen in this analysis, many RCT authors specified more than one treatment target of their exercise intervention, and these were captured by multiple outcomes. Thus, it was hypothesised that a composite outcome composed of these multiple matched outcomes might be more responsive than a single matched outcome, and more appropriate for a complex intervention such as exercise.

7.2.3 Data Analysis 2: A Matched Composite in Comparison to a Single Outcome

This further secondary analysis study aimed to evaluate whether a composite outcome consisting of multiple matched outcomes might change the results and conclusions of existing RCTs of exercise for NSLBP. Four RCT datasets were obtained that included multiple outcomes that matched the RCT authors' specified treatment targets. A total of 864 participants were included in this analysis, with two RCTs having two 'matched' primary outcomes, and two RCTs included with 'unmatched' primary outcomes. The results of this analysis demonstrated that a matched composite outcome generated a greater SMD than

a single unmatched primary outcome. Had the included RCTs used a matched composite outcome, this would have changed the results and conclusions of three of the four RCTs to be more in favour of exercise.

In the sub-group analysis of matched composites compared to unmatched primary outcomes (n= 2 trials), a matched composite outcome produced larger SMDs than an unmatched primary outcome in favour of exercise in comparison to the control arm. However, in the matched primary outcome group (n=2 trials), when the composite outcome was compared to the single, first-mentioned, matched primary outcome, there was no change in the results. A co-primary composite was then created for these two RCTs, to compare the results of a more 'targeted' composite with a matched outcome and a matched composite. In both RCTs, a co-primary composite generated a larger SMD, in favour of exercise, than a matched composite of all matched treatment targets. In one of these two RCTs, the single primary outcome still generated larger SMDs than the co-primary composite (136), whereas in the other RCT the co-primary composite generated larger SMDs than the single matched primary outcome (137), in favour of exercise.

A summary forest plot demonstrated that a single outcome generated a pooled SMD of 0.24 (95% CI -0.04, 0.53) (p=0.10)) in comparison to a composite matched outcome which generated a pooled SMD of 0.28 (95% CI 0.05, 0.51) (p=0.02)). The between-group difference was not statistically significant (SMD 0.04 (95% CI –0.13, 0.20). This difference may be because three of the composite outcomes had statistically significant results in favour of the exercise

arm, in contrast to only one of the single outcomes, although the standard error was smaller for each of the composite outcomes. Composite outcomes, or well-matched co-primary outcomes, may be preferable to a single matched primary outcome as they are more likely to cover a wider range of treatment targets more reflective of the broader benefits from a complex intervention. The results provide initial support to suggest that a composite outcome matched to the exercise treatment targets may generate larger SMDs in favour of exercise in comparison to a non-exercise control.

These secondary analyses suggested that not all treatment targets are equal, or potentially targeted to the same degree by the exercise interventions investigated in the sample of RCT datasets. Understanding the degree to which an exercise intervention changes these targets is still poorly described, and how this can be captured effectively within an intervention description and evaluation remains to be clarified. Although other research fields such as inflammatory arthritis (217,218) and cardiovascular disease (219,220) have been using composite outcomes for some time, it is not clear if these are always routinely related to the treatment targets of the interventions being tested. Furthermore, identifying outcomes to capture these treatment targets, that are both responsive and have increased domain coverage is currently difficult given that the underlying treatment targets are poorly described and understood. This led to the use of consensus workshops to try to agree and prioritise the most important treatment targets in RCTs of exercise interventions for persistent NSLBP.

7.2.4 Consensus Workshops

There is no previously published consensus on the potential treatment targets of exercise interventions for NSLBP. Despite a systematic review describing the use of physical activity for mental health problems (221), there is little to assist RCT developers and authors regarding the treatment targets of exercise interventions for persistent NSLBP apart from the poor correlations found between pain and disability with strength and flexibility targets (22,34,35). This discord is evident when one reviews the results of the systematic review performed in this thesis, in which most RCTs specified treatment targets of muscle strength (i.e. a stronger spine) (n=8) or spinal stabilisation exercises (i.e. more stable spine) (n=8) as their primary goal, despite poor evidence of relationships between with these targets and the common outcomes of pain and disability (22,34,35). This resulted in a clear need to try to develop consensus on what the most important treatment targets in exercise interventions were.

Two sequential consensus workshops were held, incorporating stakeholders of researchers developing exercise interventions for testing in RCTs, users of exercise for NSLBP, and prescribers of exercise for NSLBP. A total of 39 participants from 10 different countries participated in the workshops, and used both online and face-to-face methods to engage, identify, vote and rank the treatment targets for exercise for NSLBP. The very nature of the nominal group workshop results in a final ranked list of treatment targets. The first workshop was a national workshop with participants from across the UK and was informed by the results of the systematic review. The final prioritised treatment targets from

this workshop were presented to the participants of the international workshop, before modifying, voting and ranking according to the outputs of their discussions and actions. The overall consensus prioritised ‘improving function’, ‘reducing pain’, ‘improving quality of life’, ‘improving patient-specific goals’, ‘reducing fear of movement’ and ‘increasing physical activity’ as the most important treatment targets in RCTs of exercise for persistent NSLBP.

7.3 Discussion Points

This section discusses the overall contribution of the research within this thesis, to what extent the results might be considered expected or unanticipated, and how they compare with contemporary literature. This is followed by consideration of the strengths and limitations of the research, the implications for research and clinical practice, and the conclusions reached by this body of research.

7.3.1 Overall Contribution of this Programme of Research

The overall aim of the research was to explore whether better matching of the outcome domains used in RCTs to the treatment targets of the exercise interventions being investigated might change the results and conclusions of these trials in the field of persistent NSLBP. This exploratory programme of research found that RCTs that used an outcome that matched the specified treatment targets of their exercise intervention were more likely to have results demonstrating superiority of exercise in comparison to a non-exercise control. Composite outcomes, comprised of multiple outcomes matched to the multiple treatment targets, again were more likely to support the superiority of exercise in

comparison to a single matched or unmatched outcome. These results suggest RCT authors should specify the treatment targets of their intervention, and consider more carefully matching their primary outcome to the(se) treatment target(s) most likely to be changed by their exercise intervention.

The systematic review highlighted the wide variety in treatment targets used in RCTs of exercise for persistent NSLBP (n=31). The consensus workshops, including multiple stakeholder groups from different countries, found overall agreement in the prioritisation of 'improving function', 'reducing pain', 'improving quality of life', 'targeting patient-specific goals', 'reducing fear of movement' and 'increasing physical activity'. The logic model including these treatment targets provides a model for future trials to use when considering the development and specification of their exercise interventions. The prioritised targets from these workshop should be explored further in RCTs of exercise for NSLBP with *a priori* specified mediation (or multiple mediation) analyses, possibly by using a logic model to demonstrate the relationships between targets and outcomes.

7.3.2 Expected Results Found in this Programme of Research

7.3.2.1 **Exercise has Multiple Treatment Targets**

It was unsurprising that most RCTs identified more than one treatment target (median of 3 targets per RCT in the systematic review), given the complexity of exercise as an intervention, and thus selecting the most 'matched' primary outcome measure can be challenging. Without clear programme theory to demonstrate the most plausible and important treatment targets, how the

exercise intervention proposes to target them, and the most appropriate outcomes to capture these targets, it is difficult to move forwards as a field. Choosing the most important treatment target(s) is difficult, as with a complex intervention, such as exercise, the treatment targets may differ according to the population, the context and the phase of delivery of the intervention (52). Thus, this research prioritising the most important treatment targets may be crucial so that trial authors developing exercise interventions can consider these prioritised targets when tailoring the intervention to the most important targets that are population, context- and delivery-relevant in that RCT.

A composite outcome matched to the multiple treatment targets of the intervention may be a good starting point. However, given the heterogeneity existing in the field of exercise interventions in persistent NSLBP, it is likely that there will be multiple composite outcomes selected and created for this purpose, as in the field of cardiovascular disease (185). This research suggests that the use of logic models and programme theory to inform the development of exercise intervention may provide a helpful framework to ensure appropriate, matched, and relevant outcomes are selected that are trial specific.

7.3.2.2 **Matching of Outcomes to Treatment Targets Results in Greater SMDs**

It was not surprising that RCTs which used an outcome matched to their specified treatment targets of their intervention generated larger SMDs, and were more likely to demonstrate statistical significance in favour of the exercise arm, in comparison to the control arm. RCTs should be designed with the most

appropriate outcome(s) selected as a primary outcome to capture the change the intervention is designed to deliver. “The ultimate value of a RCT …will be directly tied to how well the selected outcome measure matches the researcher’s understanding of what he or she expects to change, to what degree it is expected to change, over what period of time this change will happen and how that change can best be identified” (83). However, it was surprising to see how few RCT authors designed their RCTs in this manner (27%).

7.3.2.3 **Multiple Outcomes Are Frequently Used in RCTs**

Unsurprisingly, ten (37%) of the included RCTs in the systematic review nominated more than one primary outcome, suggesting that the use of composite outcomes may have a role in future RCTs of exercise interventions in persistent NSLBP. RCTs should use the most important outcomes as their primary outcome (15), with no clear guidance on how many are preferable. However, it does appear that selection of the outcome components is important to ensure that the most responsive outcomes are included, ideally by ensuring a good match between the selected outcomes and the identified treatment targets of the intervention. In musculoskeletal medicine, responder indices have been developed for LBP (91,222) incorporating a benchmark on pain scores (the VAS) and physical function scores (RMDQ). Both these outcomes were most frequently reported in the systematic review findings (chapter 3) and consensus workshops, and given these are both core outcome domains, these two outcomes could be measured as a co-primary composite outcome. However, for the greatest between-arm difference to be demonstrated, these outcomes should be matched to the treatment targets of the exercise intervention. Considering

which comes first (prioritised treatment targets or outcomes) is a challenge, as it appears in much of the work to date in psoriatic arthritis (223), rheumatoid arthritis (88), cardiovascular disease (224), and LBP (78) the prioritised outcomes have already been the subject of international consensus and recommendations. The outcomes chosen for use in RCTs should be both clinically relevant and pertinent to both patients and policy-makers (187). As the outcomes are already agreed as the most important outcomes, interventions should target these, but the question remains as to whether these outcomes are the most important outcomes for patients. Simon et al. (91) suggested the use of a responder index consisting of the core outcome domains for NSLBP based on a 30% improvement in pain, no reduction in function and 30% improvement of global effect. These are the same three targets prioritised by the consensus workshops, and suggest the use of a composite responder index may be an area for further research.

7.3.3 Unanticipated Results Found in this Programme of Research

7.3.3.1 **Uncertainty of the Mechanisms Underlying Exercise for Treating Persistent NSLBP**

It was surprising that the most important treatment targets prioritised by the consensus workshops appeared to be those internationally recommended as core outcome domains for use within LBP trials (78). Core outcome domains are not intervention specific, and it might be that research active clinicians and clinical researchers (who formed the majority of the workshop participants) may be familiar with these published recommendations, without considering the

temporality or underlying targets (in this case of exercise interventions) that may precede these outcomes. These treatment target priorities agree with previous suggestions identifying potential treatment targets of exercise (49). However, it also highlights the uncertainty associated with how exercise brings about a beneficial effect on pain and physical function.

Surprisingly, despite evidence to the contrary (22,34,35), many clinicians and RCT authors appeared to select exercises for persistent NSLBP based on a biomedical model of LBP. As seen in Table 7-1, the treatment targets prioritised across the workshops differed, as did priorities of treatment targets when compared to the original list of cited treatment targets from the systematic review. Treatment targets such as ‘increased muscle strength’, were prioritised by the first, national, workshop, as well as the systematic review, and similarly documented by Stenner (225) in interviews with clinicians regarding the treatment targets underpinning exercise prescription. The number of physical performance targets (shaded grey in Table 7-1) reported in the systematic review reduces from six targets to one target in the final, international, workshop in the top ten targets. Whereas the number of psychological targets (shaded in green) increases from two to four targets in the international workshop in the top ten. This suggests a more temporal shift in clinician and researcher beliefs since our systematic review including RCTs over a wide time period (from 1993- 2018) with just over half the included RCTs dating from before 2010 (14/27).

Table 7-1: Comparison of results of treatment target priorities

Systematic Review	National Workshop	International Workshop
Reduce pain	Reduce pain	Increase function
Spinal stabilisation	Increase function	Improve quality of life
Improve muscle strength	Reduce fear of movement	Reduce pain
Improve mobility	Encourage normal movement	Meet patient-specific goals
Improve posture	Improve mobility	Reduce fear of movement
Improve self-confidence	Improve self-efficacy	Increase physical activity
Increase function	Enhance self-management skills	Improve self-efficacy
Increase physical activity	Prevent recurrence	Enhance self-management skills
Increase muscle endurance	Improve general health and well-being	Improve work capacity
Reduce stress	Improve muscle strength	Change cognitions/attitudes and beliefs

The colours represent different constructs: green are psychological targets, orange are functional, grey are physical/ performance targets, pink is pain, light blue are general health-related quality of life targets, and white are work-related targets.

It was unexpected that, over time, from the initial RCTs included in the systematic review to the two nominal group workshops, the priority treatment targets of researchers, clinicians and users of exercise for persistent NSLBP appear to have changed. The most frequently cited treatment target in the 'matched' RCTs was 'pain reduction' (71.4% of RCTs), followed by 'self-confidence' (42.9% of RCTs) and 'physical function' (42.9% of RCTs) as seen in Table 7-2. In comparison to the frequency of treatment targets reported by the systematic review (in Table 7-1), more psychological targets were specified by the 'matched' trials (5 in the 'matched' trials, 2 in all RCTs most frequently cited ten). Three of the 'matched' RCTs were published in 2018, whereas the other four were published prior to 2014 (2000-2014). This may demonstrate a temporal shift in understanding the effects and possible mechanisms of the effect of exercise, with some of the 'matched' RCT authors considering alternative mechanisms and treatment targets for the exercise intervention being tested, as demonstrated by

the wide variety of treatment targets listed in Table 7-2. This may have been influenced by the introduction of the biopsychosocial model, which was widely accepted into many areas of clinical practice in the early 2000's (226) and may have facilitated this temporal shift.

Table 7-2: Treatment targets of matched RCTs

Treatment Targets	**Citation count (n=) of 7 RCTs**
Reduce pain; modify perception pain	5
Exercise self-efficacy; increasing self-confidence Take control of their situation	3
Modify perception disability Improve disability Preventing long-term disability	3
Improve physical capability; Increasing level activity	2
Fear of physical activity Kinesiophobia	2
Posture and movement patterns	1
Stretch shortened tissues	1
Recovery	1
Improvement in the execution of proper patient lifting techniques	1
Improve muscular stabilisation	1
Improve functional capacity	1
Reduce absence from work	1
Catastrophising	1
Reduce anxiety and depression	1
Cope better	1

The colours represent different targets: green are psychological targets, orange are functional targets, light blue are physical performance targets, pink is pain and white is work-related.

7.3.3.2 Previously Tested Mediators Were Not Prioritised as Treatment Targets

It was unexpected that one of the few tested mediators of exercise for persistent NSLBP - pain catastrophising (150)- was not prioritised by this consensus exercise (final ranked position 11) but was identified as a treatment target by only one of the matched RCTs (136). As mentioned above, more recently psychological treatment targets have been tested as potential mediators in the use of exercise to change outcomes of physical function and pain in NSLBP.

Almost all of the mediators tested in previous RCTs of exercise interventions were psychological constructs, of which ‘reduced distress’, ‘reduced fear’, ‘increased self-efficacy’ and ‘reduced pain catastrophising’ have been shown to mediate the pain-disability pathway of exercise in people with back and neck pain (150,152,227). However, comparison of the final prioritised list of treatment targets from the consensus workshops with the research exploring potential mediators of exercise interventions in persistent NSLBP (please see Appendix 9.n: Summary of Mediation Analyses in LBP) shows that many of the prioritised psychological targets have been tested as mediators previously. However, the psychological targets tested (as mediators of exercise interventions for NSLBP) only explained a moderate proportion of the total effect of exercise on pain and disability outcomes (between 20-30%) (148). This suggests that there are unexplained causal mechanisms that are yet to be identified which would account for a significant proportion of the total direct effect of exercise on pain and disability outcomes (67-80%).

Mediators of exercise to date have been investigated largely by cross-sectional analyses (204,205,228). Mansell, Hill, Main, Vowles and van der Windt (229) explored mediators of psychological interventions for LBP, but these have only explained 20-30% of the pathway to changing pain-disability outcomes. Pre- and post-intervention measures (35) and two systematic reviews (22,34) comparing the correlation of physical performance measures and pain and disability outcomes have found limited, or no evidence to suggest a relationship with the identified measures and the outcomes of pain and disability. Without a clear understanding of the treatment targets and mechanisms by which exercise works

to improve pain and physical function, it is likely that RCTs of exercise interventions will continue to use these core outcomes as their primary outcomes. However, more work is required to understand the treatment targets and mechanisms of action of exercise to ensure the most appropriate and accurate outcomes are used. Potential outcomes such as composites (which may include pain and functional outcomes), may cover the wider breadth and multiple treatment targets of an exercise intervention. Detailed programme theory of how the exercise intervention is thought to have its effect on the core outcomes of pain, and physical function (if indeed these are prioritised by the RCT designers as being the most important) needs to be clearly modelled for transparent pathways and likely intermediate targets to be identified (230), such as was performed by Sherman et al. (227) and Hurley et al. (17).

As a practicing clinical academic physiotherapist, I note the discrepancy in treatment targets prioritised by clinicians and researchers, and those focused on by the exercise prescribed. A challenge for future RCTs of exercise to design the intervention in such a way that it can affect the specified treatment targets (mediators) to a greater extent so that the intervention has a greater impact on pain and physical function outcomes (231). Bronfort et al. (117) prioritised pain reduction as their primary outcome based on the views of a group of patients, but they specified that their intervention was designed to target strength and endurance. Without a clear logic model to demonstrate how the treatment targets of strength and endurance would affect a change in pain reduction, it is difficult to assume the relationship between this outcome and the treatment targets. As mentioned previously (chapter 1), these two components (strength, endurance)

have been shown to have no correlation with pain and disability outcomes (34,35). Thus, we need to ensure that if an outcome of importance is prioritised by patients, then the exercise interventions used should be designed to target those outcomes. Future research needs to ensure optimal design of exercise interventions to affect the treatment targets specified, especially if these are of a psychological nature, as exercise design to target psychological or pain components is not clearly defined (232).

7.3.4 Interpreting this research in the field of complex interventions

Exercise, as a complex intervention, can be delivered in many different forms, prescribed by different deliverers, with varying dosages, tailoring and in a variety of settings (23). The development and testing of interventions and the causal theories that underpin these interventions should be subject to rigorous development and evaluation as per guidance from the Medical Research Council (52). These updated guidelines reinforce the need for exercise to have a clear specification of treatment targets, with consideration of the context, exercise deliverer and outcomes to be used as demonstrated in the logic model on page 228. Further, the development of the exercise intervention and the prioritisation of its treatment targets should be informed by stakeholder involvement, underpinned by programme theory and logic models of the proposed theory, and continually modified, as required, in line with evaluation and in light of uncertainty surrounding the intervention deliverables (52).

7.4 Strengths and Limitations of this Research

The systematic review was robustly performed in line with PRISMA guidance (94), with a rigorous search strategy and multiple pairs of reviewers to reduce bias when extracting data. The meta-analyses performed were compared to seven different sensitivity analyses, and results were similar across all of these meta-analyses. Further, the specification of treatment targets findings are in keeping with larger review of Cochrane exercise trials (67) suggesting that these results are generalisable across all exercise trials.

A strength of the secondary data analyses is that a good proportion of the identified RCT datasets were obtained: the datasets were obtained for three of the four RCTs in the first analyses and four of the seven RCTs in the second analyses. Another strength of this work is that a sensitivity analysis was performed with ratio of means (110) to compare the results, and similar findings were obtained. These analyses were performed on a small sample of exercise RCTs in persistent NSLBP; thus, it is unknown whether these findings would be true for other complex interventions in persistent NSLBP or other conditions. However, it would make logical sense that the phenomenon of matching RCT outcomes to treatment targets, irrespective of the condition or intervention, would potentially yield greater between-arm differences in RCTs of complex interventions.

A further strength of the consensus workshops was the participation of people from ten different countries, and inclusion of stakeholders from different

backgrounds: people with a lived experience of using exercise for NSLBP, researchers involved in the development of exercise interventions to be tested in RCTs with patients with persistent NSLBP, and clinicians prescribing exercise to treat persistent NSLBP. A strength of these workshops was the inclusion of people who used exercise to manage their NSLBP, as in previous work, patients have not been included. The discrepancy between the prioritised treatment targets in the workshops, systematic review and matched RCTs' results, also highlights a limitation of the generalisability of these workshop findings: the limited number of people with a lived experience of exercise for NSLBP who were involved in the consensus workshops. Previous work (225) has highlighted the discrepancy that exists between many clinicians and patients in the act of goal setting and exercise selection; and in RCTs of exercise interventions, where standardisation of the exercise intervention is required, this becomes even more difficult to employ.

This was an exploratory programme of research that was limited by sample sizes in the systematic review, participant responses in the consensus workshops and dataset responses for the dataset analyses. However, this is the first time this type of work has been performed, and the results suggest that further work needs to be conducted to establish whether matching is important, and what type of matching is likely to be practically implementable in RCTs (primary or composite outcomes).

The data extraction performed in the systematic review may have been limited by the ability to extract exercise intervention fidelity information from the

published RCTs. This would have been very challenging to undertake, and was not possible in the remit of this thesis, but may have helped to understand the difference between exercise interventions delivered within the matched and unmatched RCTs. This would require future work to compare the protocoled exercise intervention in comparison to what was actually delivered, and would have required not only clear documentation using the TIDieR guidelines (by RCT authors) but perhaps also interviews with practitioners and patients similar to that performed by the National Exercise Referral Scheme in Wales (233). However, the work undertaken by Stenner (225) in his thesis exploring exercise prescription by clinicians and patients with persistent NSLBP, goes some way to providing insight into this.

Both secondary data analyses comprised a small sample of datasets limiting the contribution of this research. Each RCT used different exercise approaches, with varying primary follow-up points, different analysis methods, and differing treatment targets. The use of the SMD allowed comparison across different outcomes, although a limitation is that the Cochrane group had not intended it be used for this purpose: they suggested it may be more appropriate when comparing different outcome measures within the same domain (106).

7.5 Implications for Future Research

This programme of research suggests that matching the primary outcome(s) to the identified treatment targets of exercise interventions for persistent NSLBP is an important area for further research. Further analyses on a larger sample of

exercise RCTs, and also other complex interventions for persistent NSLBP is recommended. Exercise is an advocated intervention for many different clinical conditions (osteoporosis, osteoarthritis, cardiovascular disease, hypertension, diabetes etc.) and therefore future research could also explore the potential effect of matching outcomes to exercise treatment targets in these other clinical conditions.

Understanding how exercise intervention(s) create their effects is an important gap that this research has identified. It is important to be able to identify the components that will reasonably achieve change in the identified and prioritised treatment targets from the consensus workshops, but this requires more detail on the intervention itself to be provided in research reports. Given the heterogeneity in exercise interventions, delivery settings and styles, the identified components are likely to vary from RCT to RCT, both across and within exercise intervention delivery. This research programme has highlighted the way in which many exercise interventions in RCTs are poorly described (67,234), and the need for better description of interventions (64,67,160). A clear structure to achieve this would be through the use of logic models or intervention mapping (17,52). Future RCTs of exercise interventions in persistent NSLBP should clearly define their exercise treatment targets, and this research provides exploratory support for trial teams to consider better matching their outcomes to these, in order to maximise the potential of the trial to be able to detect the benefits of exercise versus control interventions.

A highlighted limitation was the small numbers of true patients involved in this programme of research. Future research should involve greater numbers of people with a lived experience of using exercise for NSLBP, in both the design of future research methodologies, as well as the interpretation thereof (235). Further consensus workshops involving greater numbers of people who have used exercise to manage and treat NSLBP will be important to identify treatment targets of importance, in contrast to that which were prioritised in this study, with a group of predominantly clinical researchers.

Future research regarding specific exercise approaches for NSLBP may involve realist evaluations of what works for whom, under which circumstances, and why (52). Involvement with stakeholders to develop clear programme theory and logic models for the proposed rationale is recommended for all specific exercise types used to treat NSLBP. These can then be tested and evaluated using RCTs to understand what modifications are required before key active ingredients can be recommended for implementation across health services. One way to support this in research may be through an amendment to the CERT to include a clear rationale for the specific exercise (67).

Some of the identified treatment targets may be mediators of exercise interventions, and thus variables on the pathway to the key outcomes such as pain, physical function and quality of life. Future RCTs of exercise for NSLBP should be designed with additional analyses in mind (such as multiple mediation analyses) to better understand the role of treatment targets in bringing about the observed benefits of exercise, as few RCTs to date have been designed with

mediation analysis in mind (233,236). However, planned mediation analyses, according to a clear rationale, should be included in advance of the RCT delivery, such as in the protocols of RCTs, to ensure sample size and other methodological challenges are accounted for (237). Further, the updated search of ongoing RCTs and protocols of exercise interventions for persistent NSLBP (Appendix 9.k) suggests that the identified mismatch of targets and outcomes continues to be a problem within this field.

The logic model developed (in chapter 3) and updated (in chapter 6) may provide a framework for future RCTs to use when developing and specifying exercise interventions for testing. Not all trials will have the funding to perform an in-depth intervention mapping approach to design the key components of their intervention, such as was performed by Hurley et al. (17). However the use of a logic model provides a transparent means to demonstrate the proposed pathways, and even a basic model such as that used by Sherman et al. (238) may provide justification and a rationale for the use of primary outcomes and intervention design. Future trials testing exercise interventions may benefit from the use of logic models to describe their underpinning rationale for their intervention, the key treatment target(s), treatment components and outcome selection.

The use of composite outcomes in future RCTs may play an important role, possibly alongside or through the use of responder indices as proposed by Simon et al. (91). Clear specification of how exercise should be designed (or prescribed) to target these outcomes of pain, function and quality of life, has not been well

documented in the NSLBP field – unlike strength (239) and endurance exercise (240). Future work should investigate and describe how exercise interventions should be tailored and implemented to affect these areas of importance to patients with NSLBP, as well as other prioritised treatment targets, such as psychological constructs.

7.6 Implications for Clinical Practice

The output of this research programme – namely consensus on the ranked importance of treatment targets for exercise - is the first step in gaining consensus across patient, clinician and researcher populations. However, these priorities may well change dependent on exercise selection, delivery and individual goals. Clinicians need to be aware of the importance of shared decision making and goal setting with patients, to be selecting treatment targets and outcomes that are relevant to the patient, and are likely to change given the intervention goals. The results of the systematic review and both secondary data analyses suggest that a clear understanding of the treatment targets of the exercise intervention delivered is important to select the most appropriate outcome measure. In my experience of clinical practice as a physiotherapist, core outcome domain questionnaires (such as the ODI, NDI and VAS) were routinely used without much further thought to the treatment targets of the prescribed exercise interventions, suggesting clinicians should also consider the mechanism through which they expect change to occur, and to capture that with the most appropriate outcome measure. Furthermore, consideration should be given to the design of the intervention in meeting the identified treatment targets. Clear

guidance is available for intervention development to target aerobic, resistance and flexibility deficits (232).

This work provides an initial list of agreed treatment targets which may be helpful for clinicians when planning and prescribing exercise interventions for persistent NSLBP. It also provides support for the use of outcome measures that are specific to the targets of the intervention delivered in clinical practice.

7.7 Summary

In summary, this exploratory programme of research supports the premise that matching the primary outcome of RCTs to the specified treatment targets of the exercise intervention in persistent NSLBP may be important. The systematic review provided initial support by showing that RCTs that matched their outcome to the treatment targets may be more likely to find statistically significant results in favour of exercise. The premise was further supported through the secondary RCT data analyses which found that the conclusions of most of the included RCTs would have changed (4 of 5, and 2 of 4) if they had better matched their trial outcomes to their identified treatment targets. Finally, a large and -group of researchers, clinicians and people with back pain experience prioritised the most important treatment targets of exercise as ‘increasing function’, ‘improving quality of life’, ‘reducing pain’, ‘targeting patient-specific goals’, ‘reducing fear of movement’ and ‘increasing physical activity’.

This programme of work has demonstrated, for the first time, the potential to evidence greater benefits of exercise for NSLBP than has been shown to date, by better matching outcomes to treatment targets in RCTs in this field. Further work to confirm these findings is needed. In addition, the consensus study on treatment targets for exercise has highlighted gaps in knowledge of treatment mechanisms and potential mediators, which warrants further study. The effect of many interventions, not only exercise in RCTs of persistent NSLBP, may have been underestimated if outcomes and treatment targets are poorly understood and, therefore, poorly matched.

8 References

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet. 2018;391(10137):2356–67.
2. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis [Internet]. 2014;0:1–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24553908
3. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management Assessment and non-invasive treatments Low back pain and sciatica in over 16s [Internet]. 2016 [cited 2017 May 8]. Available from: https://www.nice.org.uk/guidance/ng59/evidence/full-guideline-assessment-and-noninvasive-treatments-pdf-2726158003
4. Krismer M, Van Tulder M. Low back pain (non-specific) The Low Back Pain Group of the Bone and Joint Health Strategies for Europe Project. Best Pract Res Clin Rheumatol [Internet]. 2007 Feb 1 [cited 2017 Aug 30];21(1):77–91. Available from: http://www.sciencedirect.com/science/article/pii/S1521694206001082
5. Bardin LD, King P, Maher CG. Diagnostic triage for low back pain: a practical approach for primary care. MJA [Internet]. 2017 [cited 2017 Nov 8];206(6). Available from: https://www.mja.com.au/system/files/issues/206_06/10.5694mja16.00828.pdf
6. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord [Internet]. 2010 Dec 2 [cited 2017 Sep 4];11(1):144. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-11-144
7. Artus M, van der Windt D, Jordan KP, Croft PR. The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. BMC Musculoskelet Disord [Internet]. 2014 Mar 7 [cited 2017 May 8];15(1):68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24607083
8. Qaseem A, Wilt TJ, McLean RM, Forceia M. Noninvasive Treatments for Acute,

Subacute and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. Ann Intern Med. 2017;166(7):514–30.

9. Stochkendahl MJ, Kjaer P, Hartvigsen J, Kongsted A, Aaboe J, Andersen M, et al. National Clinical Guidelines for non-surgical treatment of patients with recent onset low back pain or lumbar radiculopathy. Eur Spine J [Internet]. 2018 Apr 20 [cited 2017 Jun 4];27:60–75. Available from: http://link.springer.com/10.1007/s00586-017-5099-2
10. Hong J-Y, Song K-S, Cho JH, Lee JH. An Updated Overview of Low Back Pain Management in Primary Care. Asian Spine J [Internet]. 2017 Aug [cited 2018 Feb 23];11(4):653. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28874985
11. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CWC, Chenot JF, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J [Internet]. 2018;27(11):2791–803. Available from: https://doi.org/10.1007/s00586-018-5673-2
12. Hollon S, Arean P, Craske M, Crawford K, Kivlahan D, Magnavita J, et al. Development of clinical practice guidelines. Annu Rev Clin Psychol. 2014;10:213–41.
13. Rohwer A, Pfadenhauer L, Burns J, Brereton L, Gerhardus A, Booth A, et al. Logic models help make sense of complexity in systematic reviews and health technology assessments. J Clin Epidemiol [Internet]. 2017;83:37–47. Available from: http://dx.doi.org/10.1016/j.jclinepi.2016.06.012
14. Campbell N, Murray E. Designing and evaluating complex interventions to improve health care. BMJ [Internet]. 2007;334(7591):455–9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1808182/
15. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ [Internet]. 2008 [cited 2017 May 5];1655(September):a1655. Available from: www.mrc.ac.uk/complexinterventionsguidance
16. Ball L, Ball D, Leveritt M, Ray S, Collins C, Patterson E, et al. Using logic models to enhance the methodological quality of primary health-care interventions: guidance from an intervention to promote nutrition care by general practitioners and practice nurses. Aust J Prim Health. 2016;23(1):53–60.

17. Hurley DA, Murphy LC, Hayes D, Hall AM, Toomey E, McDonough SM, et al. Using intervention mapping to develop a theory-driven, group-based complex intervention to support self-management of osteoarthritis and low back pain (SOLAS). Implement Sci [Internet]. 2016;11(1):56. Available from: http://implementationscience.biomedcentral.com/articles/10.1186/s13012-016-0418-2

18. World Health Organization. WHO | Physical activity [Internet]. WHO. World Health Organization; 2017 [cited 2017 Nov 9]. Available from: http://www.who.int/mediacentre/factsheets/fs385/en/

19. Desmeules F, Côté CH, Frémont P. Therapeutic Exercise and Orthopedic Manual Therapy for Impingement Syndrome : A Systematic Review. Clin J Sport Med [Internet]. 2003 [cited 2017 Aug 25];13:176–82. Available from: https://www.researchgate.net/profile/Francois_Desmeules2/publication/10721010_Therapeutic_Exercise_and_Orthopedic_Manual_Therapy_for_Impingement_Syndrome_A_Systematic_Review/links/00b7d53780f4c2280c000000/Therapeutic-Exercise-and-Orthopedic-Manual-Therapy

20. Naugle KM, Naugle KE, Riley JL, III. Reduced Modulation of Pain in Older Adults After Isometric and Aerobic Exercise. J Pain [Internet]. 2016 Jun [cited 2017 Nov 9];17(6):719–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26993959

21. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev [Internet]. 2017 [cited 2017 Aug 3];4(4):CD011279. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28436583

22. Steiger F, Wirth B, de Bruin ED, Mannion AF. Is a positive clinical outcome after exercise therapy for chronic non-specific low back pain contingent upon a corresponding improvement in the targeted aspect(s) of performance? A systematic review. Eur Spine J [Internet]. 2012 Apr [cited 2017 Aug 3];21(4):575–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22072093

23. Ferreira ML, Smeets RJEM, Kamper SJ, Ferreira PH, Machado LAC. Can We Explain Heterogeneity Among Randomized Clinical Trials of Exercise for Chronic Back Pain? A Meta-Regression Analysis of Randomized Controlled Trials. Phys Ther [Internet]. 2010

Oct 1 [cited 2018 Jan 5];90(10):1383–403. Available from: https://watermark.silverchair.com/ptj1383.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAbUwggGxBgkqhkiG9w0BBwagggGiMIIBngIBADCCAZcGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMgCrDvgM5ApYAOSNLAgEQgIIBaHlpA7yAUfz-gYC-pUt_yXUxgCdKWgEHs1rKKZXa7eY1Nnx

24. Owen PJ, Miller CT, Mundell NL, Verswijveren SJ, Tagliaferri SD, Brisby H, et al. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. Br J Sports Med. 2019;1–12.
25. Hayden J, van Tulder M, Tomlinson G. Systematic Review: Strategies for Using Exercise Therapy To Improve Outcomes in Chronic Low Back Pain. Ann Intern Med [Internet]. 2005 May 3 [cited 2017 Aug 11];142(9). Available from: http://www.ncbi.nlm.nih.gov/pubmed/15867410
26. Hayden J, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. In: Hayden J, editor. Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2005 [cited 2017 May 8]. p. CD000335. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16034851
27. Searle A, Spink M, Ho A, Chuter V. Exercise interventions for the treatment of chronic low back pain: a systematic review and meta-analysis of randomised controlled trials. Clin Rehabil [Internet]. 2015 Dec 13 [cited 2017 Aug 25];29(12):1155–67. Available from: http://journals.sagepub.com/doi/10.1177/0269215515570379
28. Cohen J. Statistical Power Analysis. Curr Dir Psychol Sci [Internet]. 1992 Jun 24 [cited 2018 Jan 25];1(3):98–101. Available from: http://journals.sagepub.com/doi/10.1111/1467-8721.ep10768783
29. Babatunde OO, Jordan JL, Van der Windt DA, Hill JC, Foster NE, Protheroe J. Effective treatment options for musculoskeletal pain in primary care: A systematic overview of current evidence. Fleckenstein J, editor. PLoS One [Internet]. 2017 Jun 22 [cited 2018 Feb 12];12(6):e0178621. Available from: http://dx.plos.org/10.1371/journal.pone.0178621
30. Verbunt J, Seelen H, Vlaeyen J, van de Heijden G, Heuts P, Pons K, et al. Disuse and deconditioning in chronic low back pain: concepts and hypotheses on contributing

mechanisms. Eur J Pain. 2003;7(1):9–21.

31. Bortz WMI. The disuse syndrome. West J Med [Internet]. 1984;141(November):691–4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/pmc1011199/

32. Vlaeyen JW., Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain. 2000;85:317–32.

33. Hasenbring M, Marienfeld G, Kuhlendahl D, Soyka D. Risk factors of chronicity in lumbar disc patients. A prospective investigation of biologic, psychologic, and social predictors of therapy outcome. Spine (Phila Pa 1976). 1994;19:2759–65.

34. Laird RA, Kent P, Keating JL. Modifying patterns of movement in people with low back pain -does it help? A systematic review. BMC Musculoskelet Disord [Internet]. 2012 Sep 7 [cited 2017 Aug 17];13(1):169. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22958597

35. Mannion AF, Caporaso F, Pulkovski N, Sprott H. Spine stabilisation exercises in the treatment of chronic low back pain: a good clinical outcome is not associated with improved abdominal muscle function. Eur Spine J [Internet]. 2012 Jul 24 [cited 2017 Aug 17];21(7):1301–10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3389103/pdf/586_2012_Article_2155.pdf

36. Stathopoulou G, Powers MB, Berry AC, Smits JAJ, Otto MW. Exercise Interventions for Mental Health: A Quantitative and Qualitative Review. Clin Psychol Sci Pract [Internet]. 2006 May 1 [cited 2018 Aug 20];13(2):179–93. Available from: http://doi.wiley.com/10.1111/j.1468-2850.2006.00021.x

37. Cheng YY, Kao CL, Ma HI, Hung CH, Wang CT, Liu DH, et al. SIRT1-related inhibition of pro-inflammatory responses and oxidative stress are involved in the mechanism of nonspecific low back pain relief after exercise through modulation of Toll-like receptor 4. J Biochem. 2015;158(4):299–308.

38. Nijs J, Kosek E, Van Oosterwijck J, Meeus M. Dysfunctional endogenous analgesia during exercise in patients with chronic pain: to exercise or not to exercise? Pain Physician [Internet]. 2012 Jul [cited 2018 Jul 6];15(3 Suppl):ES205-13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22786458

39. Gleeson M, Bishop NC, Stensel DJ, Lindley MR, Mastana SS, Nimmo MA. The anti-

inflammatory effects of exercise: Mechanisms and implications for the prevention and treatment of disease. Nat Rev Immunol [Internet]. 2011;11(9):607–10. Available from: http://dx.doi.org/10.1038/nri3041

40. Abdel-Hamid Hassan E, Ahmed Amin M. Pilates Exercises Influence on the Serotonin Hormone, Some Physical Variables and the Depression Degree in Battered Women. World J Sport Sci [Internet]. 2011;5(2):89–100. Available from: http://pilatesfisios.it/ricerche pdf/Pilates Exercises Influence on the Serotonin Hormone,.pdf
41. Wand BM, Parkitny L, O'Connell NE, Luomajoki H, McAuley JH, Thacker M, et al. Cortical changes in chronic low back pain: Current state of the art and implications for clinical practice. Man Ther [Internet]. 2011;16(1):15–20. Available from: http://dx.doi.org/10.1016/j.math.2010.06.008
42. Wand BM, O'Connell NE. Chronic non-specific low back pain – sub-groups or a single mechanism? BMC Musculoskelet Disord [Internet]. 2008 Dec 25 [cited 2017 Aug 21];9(1):11. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-9-11
43. Justice L, Sofka A, McGinty A. Targets, Techniques, and Treatment Contexts in Emergent Literacy Intervention. Semin Speech Lang [Internet]. 2007 Feb 6 [cited 2019 Feb 27];28(1):014–24. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-2007-967926
44. Whittle R, Mansell G, Jellema P, Windt D Van Der. Applying causal mediation methods to clinical trial data: What can we learn about why our interventions (don't) work ? Eur J Pain. 2017;21:614–22.
45. Mansell G, Kamper SJ, Kent P. Why and how back pain interventions work: What can we do to find out? Best Pract Res Clin Rheumatol [Internet]. 2013 Oct [cited 2017 Jun 30];27(5):685–97. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1521694213000776
46. Staal JB, Hlobil H, van Tulder MW, Köke AJA, Smid T, van Mechelen W. Return-to-work interventions for low back pain: a descriptive review of contents and concepts of working mechanisms. Sports Med [Internet]. 2002 [cited 2018 Jan 15];32(4):251–67. Available

from: http://www.ncbi.nlm.nih.gov/pubmed/11929354

47. Wessels T, van Tulder MM, Sigl T, Ewert T, Limm H, Stucki G. What predicts outcome in non-operative treatments of chronic low back pain? A systematic review. Eur Spine J [Internet]. 2006 Nov [cited 2017 Jun 22];15(11):1633–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16575598

48. Helmhout PH, Staal JB, Maher CG, Petersen T, Rainville J, Shaw WS. Exercise therapy and low back pain: insights and proposals to improve the design, conduct, and reporting of clinical trials. Spine (Phila Pa 1976) [Internet]. 2008 Jul 15 [cited 2018 Jan 5];33(16):1782–8. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18628711%5Cnhttp://graphics.tx.ovid.com/ovftpdfs/FPDDNCOBMECJGJ00/fs046/ovft/live/gv025/00007632/00007632-200807150-00013.pdf

49. Rainville J, Hartigan C, Martinez E, Limke J, Jouve C, Finno M. Exercise as a treatment for chronic low back pain. Spine J [Internet]. 2004 Jan 2 [cited 2019 Jul 2];4(1):106–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14749199

50. Moffett JK, Mannion AF. What is the value of physical therapies for back pain? Best Pract Res Clin Rheumatol [Internet]. 2005 Aug [cited 2018 Jan 15];19(4):623–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15949780

51. Hayden JA, Wilson MN, Stewart S, Cartwright JL, Smith AO, Riley RD, et al. Exercise treatment effect modifiers in persistent low back pain: an individual participant data meta-analysis of 3514 participants from 27 randomised controlled trials On behalf of Chronic Low Back Pain IPD Meta-Analysis Group. Br J Sport Med. 2019;0:1–16.

52. Craig P, Matthews L, Moore L, Simpson S, Skivington K. Updated guidance: developing and evaluating complex interventions [draft of updated guidance for consultation]. 2019;99. Available from: https://www.sphsu.gla.ac.uk/stakeholder-survey-2019/Full complex guidance draft for consultation v1.1 26.03.19.pdf

53. Froud R, Rajendran D, Patel S, Bright P, Bjørkli T, Eldridge S, et al. The Power of Low Back Pain Trials. Spine (Phila Pa 1976) [Internet]. 2017 Jun 1 [cited 2018 Jan 25];42(11):E680–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27792111

54. Jadad AR, Enkin MW. Randomized Controlled Trials: questions, answers and musings.

In: Randomized Controlled Trials [Internet]. Second. UK: Blackwell Publishing Ltd; 2007 [cited 2018 Feb 14]. Available from: http://doi.wiley.com/10.1002/9780470691922.fmatter

55. Costa LOP. Randomised controlled trials for complex physiotherapy interventions are perfectly possible. Br J Sports Med. 2018;52(15):950–1.

56. Østerås H, Paulsberg F, Gravare Silbernagel K. Are randomised control trials best for evaluating the effect of complex physical therapy interventions? Br J Sports Med [Internet]. 2018 Jan 3 [cited 2018 Jul 4];bjsports-2017-098456. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29298752

57. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical Evidence of Bias. JAMA [Internet]. 1995 Feb 1 [cited 2018 Jan 15];273(5):408. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1995.03520290060030

58. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ [Internet]. 2001 Jul 7 [cited 2018 Jan 15];323(7303):42–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11440947

59. Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. J Am Med Assoc [Internet]. 2001 Apr 18 [cited 2018 Jan 15];285(15):1992–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11308436

60. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. In: Kunz R, editor. Cochrane Database of Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2007 [cited 2018 Jan 15]. p. MR000012. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17443633

61. Van Tulder MW, Suttorp M, Morton S, Bouter LM, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. Spine (Phila Pa 1976) [Internet]. 2009;34(16):1685–92. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=2009438442

62. Manchikanti L, Hirsch JA, Smith HS. Evidence-based medicine, systematic reviews, and

guidelines in interventional pain management: Part 2: Randomized controlled trials. Pain Physician [Internet]. 2008 [cited 2018 Jan 5];11(6):717–73. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-64049119807&partnerID=40&md5=c660f16f6ba21cc8817ac83606a60db2

63. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials [Internet]. 2010 Jan [cited 2015 Apr 30];11:32. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2857832&tool=pmcentrez&rendertype=abstract

64. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ [Internet]. 2014 Mar 7 [cited 2017 Jul 7];348:1687. Available from: http://www.bmj.com/content/348/bmj.g1687

65. Slade SC, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on Exercise Reporting Template (CERT): Modified Delphi Study. Phys Ther [Internet]. 2016 Oct 1 [cited 2018 Jan 18];96(10):1514–24. Available from: https://academic.oup.com/ptj/article-lookup/doi/10.2522/ptj.20150668

66. Kent P, O'Sullivan PB, Keating J, Slade SC. Evidence-based exercise prescription is facilitated by the Consensus on Exercise Reporting Template (CERT). Br J Sports Med [Internet]. 2018 Feb 1;52(3):147 LP – 148. Available from: http://bjsm.bmj.com/content/52/3/147.abstract

67. Wood L, Ogilvie R, Hayden JA. Specifying the treatment targets of exercise interventions: do we? Br J Sports Med. 2020;0:1–2.

68. Gianola S, Frigerio P, Agostini M, Bolotta R, Castellini G, Corbetta D, et al. Completeness of Outcomes Description Reported in Low Back Pain Rehabilitation Interventions: A Survey of 185 Randomized Trials. Physiother Canada [Internet]. 2016 Aug [cited 2017 May 16];68(3):267–74. Available from: http://utpjournals.press/doi/10.3138/ptc.2015-30

69. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino M-A, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. J Clin

Epidemiol [Internet]. 2014 Jul [cited 2018 Aug 22];67(7):745–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24582946

70. Froud R, Patel S, Rajendran D, Bright P, Bj?rkli T, Buchbinder R, et al. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: Respice, adspice, et prospice. PLoS One. 2016;11(10):1–16.

71. Chiarotto A, Ostelo RW, Turk DC, Buchbinder R, Boers M. Core outcome sets for research and clinical practice. Brazilian J Phys Ther [Internet]. 2017 [cited 2017 May 24];21(2):77–84. Available from: http://www.sciencedirect.com/science/article/pii/S141335551730028X

72. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JPT, Sterne JAC, et al. Evidence for the Selective Reporting of Analyses and Discrepancies in Clinical Trials: A Systematic Review of Cohort Studies of Clinical Trials. Fugh-Berman AJ, editor. PLoS Med [Internet]. 2014 Jun 24 [cited 2017 Nov 23];11(6):e1001666. Available from: http://dx.plos.org/10.1371/journal.pmed.1001666

73. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. Boutron I, editor. PLoS One [Internet]. 2013 Jul 5 [cited 2017 Nov 23];8(7):e66844. Available from: http://dx.plos.org/10.1371/journal.pone.0066844

74. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976) [Internet]. 1998 Sep 15 [cited 2017 May 11];23(18):2003–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9779535

75. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain [Internet]. 2014 Sep [cited 2017 May 11];15(6):569–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26388962

76. Dworkin RH, Turk DC, Farrar JT, Haythornwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain [Internet]. 2005 [cited 2017 May 11];113:9–19. Available from:

http://s3.amazonaws.com/academia.edu.documents/44099162/Dworkin_RH_Turk_DC_Farrar_JT_et_al._Core20160325-13448-1a0odwl.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1494505052&Signature=ksGYx4faLjElu8SyFAz1oQSXmeU%253D&response-content-disposition=inlin

77. Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta Orthop [Internet]. 2015 Sep 3 [cited 2017 May 16];86(4):523–33. Available from: http://www.ichom.org/wp-content/uploads/2013/10/Clement-Acta-Orthop-2015_Propised-set-metrics-for-standardized-outcomes-in-mt-of-LBP-1.pdf

78. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J [Internet]. 2015 Jun 5 [cited 2017 May 16];24(6):1127–42. Available from: http://link.springer.com/10.1007/s00586-015-3892-3

79. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. Trials [Internet]. 2017 Jun 20 [cited 2017 Nov 27];18(S3):280. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/s13063-017-1978-4

80. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev [Internet]. 2016;5(1):11. Available from: http://www.systematicreviewsjournal.com/content/5/1/11

81. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials [Internet]. 2007 Nov 26 [cited 2017 May 11];8:38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18039364

82. Chiarotto A, Terwee CB, Ostelo RW. Choosing the right outcome measurement instruments for patients with low back pain. Best Pract Res Clin Rheumatol [Internet]. 2016 Dec 1 [cited 2018 Jan 25];30(6):1003–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29103546

83. Coster WJ. Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies MeSH TERMS clinical trials as topic decision making guidelines as

topic outcome assessment (health care) treatment outcome. Am J Occup Ther [Internet]. 2013 [cited 2017 May 18];67:162–70. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628620/pdf/162.pdf

84. van Tulder M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance: trials on exercise therapy for chronic low back pain as example. Spine (Phila Pa 1976) [Internet]. 2007 Jul 15 [cited 2018 Jan 5];32(16):1785–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17632400

85. Watt H, Harris M, Noyes J, Whitaker R, Hoare Z, Edwards RT, et al. Development of a composite outcome score for a complex intervention - measuring the impact of Community Health Workers. Trials. 2015;16(1):107.

86. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, et al. Evaluating Common Outcomes for Measuring Treatment Success for Chronic Low Back Pain. Spine (Phila Pa 1976) [Internet]. 2011 Oct 1 [cited 2017 May 25];36(21):S54–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21952190

87. Gilron I, Jensen MP. Clinical Trial Methodology of Pain Treatment Studies. Reg Anesth Pain Med [Internet]. 2011 Jul [cited 2017 Dec 7];36(4):374–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21610560

88. Smolen JS, Aletaha D, Bijlsma JWJ, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis [Internet]. 2010 Apr 1 [cited 2019 May 30];69(4):631–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20215140

89. Anderson J, Caplan L, Yazdany J, Robbins ML, Neogi T, Michaud K, et al. Rheumatoid Arthritis Disease Activity Measures: American College of Rheumatology Recommendations for Use in Clinical Practice. 2012 [cited 2019 May 30]; Available from: http://onlinelibrary.

90. Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, et al. Cardiovascular Outcomes Trials in Type 2 Diabetes: Where Do We Go From Here? Ref lections From a Diabetes Care Editors' Expert Forum. Diabetes Care [Internet]. 2018 [cited 2019 Jan 24];41. Available from: http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dci17-0057/-

/DC1.http://www.diabetesjournals.org/content/diabetes-core-update-podcasts.

91. Simon LS, Evans C, Katz N, Bombardier C, West C, Robbins J, et al. Preliminary development of a responder index for chronic low back pain. J Rheumatol [Internet]. 2007 Jun [cited 2017 May 11];34(6):1386–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17552065

92. O'Keeffe M, Purtill H, Kennedy N, Conneely M, Hurley J, O'Sullivan P, et al. Comparative Effectiveness of Conservative Interventions for Nonspecific Chronic Spinal Pain: Physical, Behavioral/Psychologically Informed, or Combined? A Systematic Review and Meta-Analysis. J Pain [Internet]. 2016;17(7):755–74. Available from: http://dx.doi.org/10.1016/j.jpain.2016.01.473

93. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343(7829):1–9.

94. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Ioannidis JP a, Clarke M, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions. Ann Intern Med [Internet]. 2009 Jul 21 [cited 2017 Aug 17];151(4):W65–94. Available from: http://dx.plos.org/10.1371/journal.pmed.1000100

95. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials [Internet]. 2012 Dec 6 [cited 2017 Nov 27];13(1):132. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-132

96. Cohen J. A power primer. Psychol Bull [Internet]. 1992 Jul [cited 2018 Jan 25];112(1):155–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19565683

97. Rubinstein SM, van Eekelen R, Oosterhuis T, de Boer MR, Ostelo RWJG, van Tulder MW. The Risk of Bias and Sample Size of Trials of Spinal Manipulative Therapy for Low Back and Neck Pain: Analysis and Recommendations. J Manipulative Physiol Ther [Internet]. 2014 Oct 1 [cited 2018 Apr 26];37(8):523–41. Available from: https://www.sciencedirect.com/science/article/pii/S0161475414001237?via%3Dihub

98. Gupta SK. Intention-to-treat concept: A review. Perspect Clin Res [Internet]. 2011 Jul

[cited 2018 Aug 9];2(3):109–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21897887

99. Joseph R, Sim J, Ogollah R, Lewis M. An empirical evaluation of the impact of missing data on treatment effect. Trials [Internet]. 2015 Dec 16 [cited 2018 Aug 10];16(S2):P136. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-16-S2-P136

100. Furlan AD, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976) [Internet]. 2009 Aug 15 [cited 2015 Mar 3];34(18):1929–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19680101

101. Higgins J, Deeks J (editors). Selecting studies and collecting data. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011) [Internet]. 5.1.0. Cochrane Reviewers; 2011. Available from: . Available from www.handbook.cochrane.orge Cochrane

102. Allmark P, Baxter S, Goyder E, Guillaume L, Crofton-Martin G. Assessing the health benefits of advice services: Using research evidence and logic model methods to explore complex pathways. Heal Soc Care Community. 2013;21(1):59–68.

103. Diaz-Ordaz K, Froud R, Sheehan B, Eldridge S. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. BMC Med Res Methodol [Internet]. 2013 Dec 22 [cited 2018 Jan 18];13(1):127. Available from: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-127

104. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. BMJ [Internet]. 2008 Apr 19 [cited 2018 Jan 18];336(7649):876–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18364360

105. Froud R, Bjørkli T, Bright P, Rajendran D, Buchbinder R, Underwood M, et al. The effect of journal impact factor, reporting conflicts, and reporting funding sources, on standardized effect sizes in back pain trials: a systematic review and meta-regression. BMC Musculoskelet Disord [Internet]. 2015 Nov 30 [cited 2017 Aug 3];16:370. Available

from: http://www.ncbi.nlm.nih.gov/pubmed/26620449

106. Deeks JJ, Higgins JP, Altman DG. Chapter 9 Analysing data and undertaking meta-analyses. In: Jonathan J Deeks JPH and DGA, editor. Cochrane Handbook for Systematic Reviews Version 51 [Internet]. 2011 [cited 2018 May 23]. Available from: http://handbook-5-1.cochrane.org/chapter_9/9_analysing_data_and_undertaking_meta_analyses.htm

107. Altman DG. Practical statistics for medical research. Chapman and Hall; 1991. 611 p.

108. Dent L, Raftery J. Treatment success in pragmatic randomised controlled trials: A review of trials funded by the UK Health Technology Assessment programme. In: Clinical Trials Methodology Conference 2011 Bristol United KingdomConference Start: 20111004 Conference End: 20111005. 2011. p. 109.

109. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ [Internet]. 2003 Sep 6 [cited 2018 Aug 14];327(7414):557–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12958120

110. Fu R, Vandermeer BW, Shamliyan TA, O'Neil ME, Yazdi F, Fox SH, et al. Handling Continuous Outcomes in Quantitative Synthesis [Internet]. Handling Continuous Outcomes in Quantitative Synthesis. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Oregon; 2013. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24006546

111. Friedrich JO, Adhikari NKJ, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study. BMC Med Res Methodol. 2008;8.

112. Doi SAR, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. Contemp Clin Trials [Internet]. 2015;45:130–8. Available from: http://dx.doi.org/10.1016/j.cct.2015.05.009

113. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting Change Scores for Pain and Functional Status in Low Back Pain. Spine (Phila Pa 1976) [Internet]. 2008 Jan 1 [cited 2018 Jan 22];33(1):90–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18165753

114. Tanner-Smith EE, Grant S. Meta-Analysis of Complex Interventions. Annu Rev Public Health [Internet]. 2018;391617(1):1–16. Available from: https://doi.org/10.1146/annurev-publhealth-040617-014112

115. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ [Internet]. 2015 Mar 19 [cited 2017 Jun 19];350(19 6):h1258–h1258. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.h1258

116. Albaladejo C, Kovacs FM, Royuela A, del Pino R, Zamora J, Network SBPR. The Efficacy of a Short Education Program and a Short Physiotherapy Program for Treating Low Back Pain in Primary Care: A Cluster Randomized Trial. Spine (Phila Pa 1976). 2010;35(5):483–96.

117. Bronfort G, Maiers MJ, Evans RL, Schulz CA, Bracha Y, Svendsen KH, et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: A randomized clinical trial. Spine J [Internet]. 2011;11(7):585–98. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=104576294&site=ehost-live NS -

118. Cambron JA, Gudavalli MR, Hedeker D, McGregor M, Jedlicka J, Keenum M, et al. One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. J Altern Complement Med. 2006;12(7):659–68.

119. Cecchi F, Molino-Lova R, Chiti M, Pasquini G, Paperini A, Conti A, et al. Spinal manipulation compared with back school and with individually delivered physiotherapy for the treatment of chronic low back pain: a randomized trial with one-year follow-up. Clin Rehabil [Internet]. 2010;24(1):26–36. Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/194/CN-00742194/frame.html

120. Chen HM, Wang HH, Chen CH, Hu HM. Effectiveness of a stretching exercise program on low back pain and exercise self-efficacy among nurses in Taiwan: A randomized clinical trial. Pain Manag Nurs [Internet]. 2014;15(1):283–91. Available from: http://dx.doi.org/10.1016/j.pmn.2012.10.003

121. Chown M, Whittamore L, Rush M, Allan S, Stott D, Archer M. A prospective study of patients with chronic back pain randomised to group exercise, physiotherapy or osteopathy. Physiotherapy. 2008;94(1):21–8.

122. Costa LOP, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, et al. Motor control exercise for chronic low back pain: a randomized placebo-controlled trial. Phys Ther. 2009;89(12):1275–86.

123. Díaz Arribas MJ, Ramos Sánchez M, Pardo Hervás P, López Chicharro J, Angulo Carreré T, Ortega Molina P, et al. Effectiveness of the physical therapy Godelive Denys-Struyf method for nonspecific low back pain: primary care randomized control trial. Spine (Phila Pa 1976). 2009;34(15):1529–38.

124. Ferreira ML, Ferreira PH, Latimer J, Herbert RD, Hodges PW, Jennings MD, et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. Pain [Internet]. 2007 Sep [cited 2018 Mar 1];131(1):31–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17250965

125. Garcia AN, Costa L da CMLOP, Hancock MJ, De Souza FS, Gomes GVF de O, Almeida MO De, et al. McKenzie Method of Mechanical Diagnosis and Therapy was slightly more effective than placebo for pain, but not for disability, in patients with chronic non-specific low back pain: A randomised placebo controlled trial with short and longer term follow-up. Br J Sports Med [Internet]. 2018;52(9):594–8. Available from: http://bjsm.bmj.com/lookup/doi/10.1136/bjsports-2016-097327

126. Goldby LJ, Moore AP, Doust J, Trew ME. A Randomized Controlled Trial Investigating the Efficiency of Musculoskeletal Physiotherapy on Chronic Low Back Disorder. Spine (Phila Pa 1976) [Internet]. 2006;31(10):1083–93. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00007632-200605010-00004

127. Groessl EJ, Liu L, Chang DG, Wetherell JL, Bormann JE, Atkinson JH, et al. Yoga for Military Veterans with Chronic Low Back Pain: A Randomized Clinical Trial. Am J Prev Med. 2017;53(5):599–608.

128. Hansen F, Bendix T, Skov P, Jensen C, Kristensen J, Krohn L. Intensive, dynamic back-muscle exercises, conventional physiotherapy, or placebo-control treatment of low-back

pain: a randomized, observer-blind trial. Spine (Phila Pa 1976) [Internet]. 1993;18:98–108. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=amed&AN=9119235 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:ameddb&id=pmid:&id=doi:&issn=0362-2436&isbn=&volume=18&issue=1&spage=98&pages=98-108&date=1993&title=Spine&atit

129. Hall AM, Maher CG, Lam P, Ferreira M, Latimer J. Tai chi exercise for treatment of pain and disability in people with persistent low back pain: A randomized controlled trial. Arthritis Care Res (Hoboken) [Internet]. 2011;63(11):1576–83. Available from: http://doi.wiley.com/10.1002/acr.20594

130. Harris A, Moe TF, Eriksen HR, Tangen T, Lie SA, Tveito TH, et al. Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (The CINS trial). Eur J Pain (United Kingdom) [Internet]. 2017;21(8):1397–407. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emex&AN=617790058 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:&id=doi:10.1002%2Fejp.1041&issn=1090-3801&isbn=&volume=21&issue=8&spage=1397&pages=1397-1407&d

131. Hildebrandt VH, Roper KI, Van den B, Douwes M, Van den Heuvel SG, Van Buuren S. Cesar therapy is temporarily more effective than a standard treatment from the general practitioner in patients with chronic aspecific lower back pain; randomized, controlled and blinded study with a I year follow-up. Ned Tijdschr Geneeskd [Internet]. 2000;144(47 PG-2258–2264):2258–64. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=30845880 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:11109471&id=doi:&issn=0028-2162&isbn=&volume=144&issue=47&spage=2258&pages=2258-2264&date=2000

132. Jans MP, de Korte EM, Heinrich J, Hildebrandt VH. Intermittent follow-up treatment with Cesar exercise therapy in patients with subacute or chronic aspecific low back pain: results of a randomized, controlled trial with a 1.5-year follow-up. Ned Tijdschr Voor Fysiother [Internet]. 2006;116(5 PG-111–116):111–6. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106215682&site=ehost-live NS -

133. Járomi M, Kukla A, Szilágyi B, Simon-Ugron Á, Bobály VK, Makai A, et al. Back School programme for nurses has reduced low back pain levels: A randomised controlled trial. J Clin Nurs. 2018;27(5–6):e895–902.

134. Johnson RE, Jones GT, Wiles NJ, Chaddock C, Potter RG, Roberts C, et al. Active Exercise, Education, and Cognitive Behavioral Therapy for Persistent Disabling Low Back Pain: A Randomized Controlled Trial. Spine (Phila Pa 1976) [Internet]. 2007 [cited 2015 Jan 20];32(15):1578–85. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=47051839 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:17621203&id=doi:10.1097%2FBRS.0b013e318074f890&issn=0362-2436&isbn=&volume=32&issue=15&spage=

135. Maul I, Läubli T, Oliveri M, Krueger H. Long-term effects of supervised physical training in secondary prevention of low back pain. Eur Spine J [Internet]. 2005 Aug [cited 2017 Aug 31];14(6):599–611. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15714351

136. Miyamoto GC, Franco KFM, van Dongen JM, Franco YRDS, de Oliveira NTB, Amaral DDV, et al. Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation. Br J Sports Med [Internet]. 2018;52:859–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29525763

137. Moffett JK, Jackson DA, Gardiner ED, Torgerson DJ, Coulton S, Eaton S, et al. Randomized trial of two physiotherapy interventions for primary care neck and back pain patients: “McKenzie” vs brief physiotherapy pain management. Rheumatology. 2006;45(12):1514–21.

138. Russell I, Underwood M, Brealey S, Burton K, Coulton S, Farrin A, et al. United Kingdom

back pain exercise and manipulation (UK BEAM) randomised trial: Effectiveness of physical treatments for back pain in primary care. BMJ. 2004;329(7479):1377–81.

139. Saper RB, Lemaster C, Delitto A, Sherman KJ, Herman PM, Sadikova E, et al. Yoga, Physical Therapy, or Education for Chronic Low Back Pain: a Randomized Noninferiority Trial. Ann Intern Med [Internet]. 2017;167(2):85–94. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/559/CN-01394559/frame.html NS -

140. Shirado O, Doi T, Akai M, Hoshino Y, Fujino K, Hayashi K, et al. Multicenter randomized controlled trial to evaluate the effect of home-based exercise on patients with chronic low back pain: the Japan low back pain exercise therapy study. Spine (Phila Pa 1976) [Internet]. 2010;35(17):E811-9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20628332%5Cnhttp://graphics.tx.ovid.com.ezproxy.library.uwa.edu.au/ovftpdfs/FPDDNCOBHACOMD00/fs046/ovft/live/gv025/00007632/00007632-201008010-00015.pdf

141. Storrø S, Moen J, Svebak S. Effects on sick-leave of a multidisciplinary rehabilitation programme for chronic low back, neck or shoulder pain: Comparison with usual treatment. J Rehabil Med [Internet]. 2004;36(1):12–6. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106761959&site=ehost-live NS -

142. Tilbrook HEHE, Cox H, Hewitt CE, Kang'ombe AR, Chuang LH, Jayakody S, et al. Yoga for Chronic Low Back Pain. Ann Intern Med [Internet]. 2011 [cited 2015 Aug 11];155(9):569–78. Available from: http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=3&SID=Y1RhKDqaBkBvkTquAy4&page=1&doc=23

143. Hayden JA, van Tulder MW, Tomlinson G. Systematic Review: Strategies for Using Exercise Therapy To Improve Outcomes in Chronic Low Back Pain. Ann Intern Med [Internet]. 2005 May 3 [cited 2017 Aug 11];142(9). Available from: http://www.ncbi.nlm.nih.gov/pubmed/15867410

144. Groessl EJ, Schmalzl L, Maiya M, Liu L, Goodman D, Chang DG, et al. Yoga for veterans with chronic low back pain: Design and methods of a randomized clinical trial.

Contemp Clin Trials [Internet]. 2016;48(March):110–8. Available from: http://dx.doi.org/10.1016/j.cct.2016.04.006

145. Knapen J, Vancampfort D, Moriën Y, Marchal Y. Exercise therapy improves both mental and physical health in patients with major depression. Disabil Rehabil [Internet]. 2015 Jul 31 [cited 2018 Aug 17];37(16):1490–5. Available from: http://www.tandfonline.com/doi/full/10.3109/09638288.2014.972579

146. Helmhout PH, Staal JB, Maher CG, Petersen T, Rainville J, Shaw WS. Exercise Therapy and Low Back Pain. Spine (Phila Pa 1976) [Internet]. 2008 Jul 15 [cited 2018 Jan 5];33(16):1782–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18628711

147. Falla D, Hodges PW. Individualized Exercise Interventions for Spinal Pain. Exerc Sport Sci Rev [Internet]. 2017 Apr [cited 2018 Aug 15];45(2):105–15. Available from: http://insights.ovid.com/crossref?an=00003677-201704000-00008

148. Lee H, Mansell G, McAuley JH, Kamper SJ. Causal mechanisms in the clinical course and treatment of back pain. Best Pract Res Clin Rheumatol [Internet]. 2017;1–10. Available from: http://dx.doi.org/10.1016/j.berh.2017.04.001

149. Hall AM, Kamper SJ, Emsley R, Maher CG. Complementary Therapies in Medicine Does pain-catastrophising mediate the effect of tai chi on treatment outcomes for people with low back pain ? Complement Ther Med [Internet]. 2016;25:61–6. Available from: http://dx.doi.org/10.1016/j.ctim.2015.12.013

150. Smeets RJEM, Vlaeyen JWS, Kester ADM, Knottnerus JA. Reduction of Pain Catastrophizing Mediates the Outcome of Both Physical and Cognitive-Behavioral Treatment in Chronic Low Back Pain. J Pain [Internet]. 2006 Apr [cited 2017 Nov 8];7(4):261–71. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=psyh&AN=2006-05129-005&site=ehost-live NS -

151. Focht BC, Rejeski WJ, Ambrosius WT, Katula JA, Messier SP. Exercise, self-efficacy, and mobility performance in overweight and obese older adults with knee osteoarthritis. Arthritis Rheum [Internet]. 2005 Oct 15 [cited 2018 Jun 13];53(5):659–65. Available from: http://doi.wiley.com/10.1002/art.21466

152. Lee H, Hübscher M, Moseley GL, Kamper SJ, Traeger AC, Mansell G, et al. How does

pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. Pain. 2015;(MARCH).

153. Turner J, Anderson M, Balderson B, Cook A, Sherman K, Cherkin D. Mindfulness-based stress reduction and cognitive-behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial. Pain. 2016;157(11):2434–44.

154. Day MA, Ward C, Ehde DM, Thorn BE, Burns J, Barnier A, et al. A pilot randomized controlled trial comparing mindfulness meditation, cognitive therapy, and mindfulness-based cognitive therapy for chronic low back pain. Pain Med. 2019;20(11):2134–48.

155. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine (Phila Pa 1976) [Internet]. 2000 Dec 15 [cited 2017 May 11];25(24):3100–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11124724

156. Sterne J, Savovic J, Page M, Elbers R, Blencowe N, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ [Internet]. 2019;(July):1–24. Available from: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials

157. Stein DS. Prevention of spinal degenerative diseases as an intervention of work-related health support for longshoremen. Prävention und Rehabil [Internet]. 2004;16(2 PG-72–80):72–80. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=amed&AN=0064938 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:ameddb&id=pmid:&id=doi:&issn=0937-552X&isbn=&volume=16&issue=2&spage=72&pages=72-80&date=2004&title=Pravention+

158. Candy B, Vickerstaff V, Jones L, King M. Description of complex interventions: analysis of changes in reporting in randomised trials since 2002. Trials [Internet]. 2018 Dec 22 [cited 2018 Jun 5];19(1):110. Available from: https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-2503-0

159. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and Processes of the

CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. Ann Intern Med [Internet]. 2008 Feb 19 [cited 2018 Jun 18];148(4):W-60. Available from: http://annals.org/article.aspx?doi=10.7326/0003-4819-148-4-200802190-00008-w1

160. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? BMJ [Internet]. 2008 Jun 28 [cited 2018 Jun 5];336(7659):1472–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18583680

161. Schroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments : a review of published randomised controlled trials. BMJ Open. 2012;2:e001978.

162. Higgins JP, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011). 2011.

163. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? BMC Med Res Methodol [Internet]. 2014 Feb 21 [cited 2018 Jun 11];14:30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24559167

164. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Internet]. The Cochrane Collaboration; 2011. Available from: www.handbook.cochrane.org.

165. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods. 2017;8(February):290–302.

166. Parkes M, Callaghan M, Tive L, Lunt M, Felson D. Responsiveness of Single versus Composite Measures of Pain in Knee Osteoarthritis. J Rheumatol. 2018;

167. Homack SR. Understanding what ANOVA Post Hoc Tests Are, Really. In: Annual Meeting of the Southwest Educational Research Association [Internet]. New Orleans, LA; 2001 [cited 2019 Jun 12]. Available from: https://files.eric.ed.gov/fulltext/ED449222.pdf

168. Higgins JPT, Deeks J (editors). Chapter 7: Selecting studies and collecting data. In: Cochrane handbook for systematic reviews of interventions Version 510. 2011. p. 1–28.

169. Hesselmann G. Applying Linear Mixed Effects Models (LMMs) in Within-Participant

Designs With Subjective Trial-Based Assessments of Awareness—a Caveat. Front Psychol [Internet]. 2018 May 25 [cited 2019 Apr 9];9:788. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00788/full

170. Hall A, Richmond H, Copsey B, Hansen Z, Williamson E, Jones G, et al. Physiotherapist-delivered cognitive-behavioural interventions are effective for low back pain, but can they be replicated in clinical practice? A systematic review. Disabil Rehabil [Internet]. 2018 Jan 2 [cited 2019 Apr 9];40(1):1–9. Available from: https://www.tandfonline.com/doi/full/10.1080/09638288.2016.1236155

171. Peres-Neto PR. Conducting Multiple Ecological Inferences Revisited. Mar Ecol Prog Ser. 1999;176:303–6.

172. Cheng HG, Phillips MR. Secondary analysis of existing data: opportunities and implementation. Shanghai Arch Psychiatry. 2014;26(6):371–5.

173. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. BMJ. 2010;341(7769):381.

174. Parkes MJ, Callaghan MJ, Tive L, Lunt M, Felson DT. Responsiveness of single versus composite measures of pain in knee osteoarthritis. J Rheumatol. 2018;45(9):1308–15.

175. Song M-K, Lin F-C, Ward S, Fine J, Hill C. Composite Variables: When and How. Nurs Res. 2013;62(1):45–9.

176. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 1988. 567 p.

177. Schunemann H, Oxman A, Higgins J, Vist G, Glasziou P, Guyatt G. 11 Presenting results and Summary of findings tables [Internet]. 5.1. Higgins J, Green S, editors. Cochrane Statistical methods Group; 2011 [cited 2019 May 30]. Available from: https://handbook-5-1.cochrane.org/chapter_11/11_presenting_results_and_summary_of_findings_tables.htm

178. Halliday MH, Garcia AN, Amorim AB, Machado GC, Hayden JA, Pappas E, et al. Treatment Effect Sizes of Mechanical Diagnosis and Therapy for Pain and Disability in Patients With Low Back Pain: A Systematic Review. J Orthop Sport Phys Ther [Internet]. 2019 Apr [cited 2019 Jun 7];49(4):219–29. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/30759358

179. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol [Internet]. 2010;63(7):737–45. Available from: http://dx.doi.org/10.1016/j.jclinepi.2010.02.006

180. Armstrong PW, Westerhout CM. Composite End Points in Clinical Research: A Time for Reappraisal. Circulation. 2017;135(23):2299–307.

181. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. Am J Obstet Gynecol. 2007;196(2):119.e1-119.e6.

182. Hall AM, Maher CG, Latimer J, Ferreira ML, Costa LOP. The patient-specific functional scale is more responsive than the Roland Morris disability questionnaire when activity limitation is low. Eur Spine J [Internet]. 2011 Jan [cited 2018 Dec 13];20(1):79–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20628767

183. Chiarotto A. A Core Outcome Measurement Set for Low Back Pain. Amsterdam, Netherlands: Vrije Universiteit; 2018.

184. Turk DC, Dworkin RH, McDermott MP, Bellamy N, Burke LB, Chandler JM, et al. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. Pain [Internet]. 2008;139(3):485–93. Available from: http://dx.doi.org/10.1016/j.pain.2008.06.025

185. Goldfarb M, Drudi L, Almohammadi M, Langlois Y, Noiseux N, Perrault L, et al. Outcome reporting in cardiac surgery trials: Systematic review and critical appraisal. J Am Heart Assoc. 2015;4(8):1–9.

186. Vaanholt MCW, Kok MM, von Birgelen C, Weernink MGM, van Til JA. Are component endpoints equal? A preference study into the practice of composite endpoints in clinical trials. Heal Expect. 2018;21(6):1046–55.

187. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials. 2017;18(1):1–7.

188. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. BMC Med Res Methodol. 2003;3:1–9.

189. Freemantle N, Calvert MJ. Interpreting composite outcomes in trials. BMJ [Internet]. 2010 Aug 18 [cited 2018 Nov 28];341:c3529. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20719822

190. Chiarotto A, Terwee CB, Deyo RA, Boers M, Lin C-WC, Buchbinder R, et al. A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. Trials [Internet]. 2014 Dec 26 [cited 2018 Mar 8];15(1):511. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-15-511

191. Harman NL, Bruce IA, Callery P, Tierney S, Sharif MO, O'Brien K, et al. MOMENT – Management of Otitis Media with Effusion in Cleft Palate: protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. Trials [Internet]. 2013 Mar 12 [cited 2018 Mar 8];14(1):70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23497540

192. MacLennan S, Bekema HJ, Williamson PR, Campbell MK, Stewart F, MacLennan SJ, et al. A core outcome set for localised prostate cancer effectiveness trials: protocol for a systematic review of the literature and stakeholder involvement through interviews and a Delphi survey. Trials [Internet]. 2015 Dec 4 [cited 2018 Mar 8];16(1):76. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/s13063-015-0598-0

193. Waggoner J, Carline JD, Durning SJ. Is There a Consensus on Consensus Methodology? Descriptions and Recommendations for Future Consensus Research. Acad Med [Internet]. 2016;91(5):663–8. Available from: http://insights.ovid.com/crossref?an=00001888-201605000-00022

194. Van de Ven AH, Delbecq AL. The nominal group as a research instrument for exploratory health studies. Am J Public Health [Internet]. 1972 Mar [cited 2018 Oct 29];62(3):337–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/5011164

195. Harvey N, Holmes CA. Nominal group technique: An effective method for obtaining group consensus. Int J Nurs Pract [Internet]. 2012 Apr [cited 2018 Nov 1];18(2):188–94. Available from: http://doi.wiley.com/10.1111/j.1440-172X.2012.02017.x

196. Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The Nominal Group Technique: A Research Tool for General Practice? [Internet]. Vol. 10, Family Practice © Oxford

University Press. 1993 [cited 2018 Nov 1]. Available from: https://academic.oup.com/fampra/article-abstract/10/1/76/577561

197. Mc Sharry J, Fredrix M, Hynes L, Byrne M. Prioritising target behaviours for research in diabetes: Using the nominal group technique to achieve consensus from key stakeholders. Res Involv Engagem [Internet]. 2016 Dec 6 [cited 2018 Nov 1];2(1):14. Available from: http://researchinvolvement.biomedcentral.com/articles/10.1186/s40900-016-0028-9

198. Potter M, Gordon S, Hamer P. The Nominal Group Technique: A useful consensus methodology in physiotherapy research. New Zeal J Physiother [Internet]. 2004;32(3):126–30. Available from: http://physiotherapy.org.nz/assets/Professional-dev/Journal/2004-November/2004NovHamner.pdf

199. Black N. Consensus Development Methods. In: Pope C, Mays N, editors. Qualitative Research in Health Care, Third Edition. 2006. p. Chapter 12.

200. Cantrill JA, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. Int J Pharm Pract [Internet]. 1996 Jun 1 [cited 2018 Nov 1];4(2):67–74. Available from: http://doi.wiley.com/10.1111/j.2042-7174.1996.tb00844.x

201. Haesler E, Carville K, Haesler P. Priority issues for pressure injury research: An Australian consensus study. Res Nurs Heal [Internet]. 2018 [cited 2018 Nov 1];41:355–68. Available from: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nur.21878

202. Woby SR, Watson PJ, Roach NK, Urmston M. Are changes in fear-avoidance beliefs, catastrophizing, and appraisals of control, predictive of changes in chronic low back pain and disability? Eur J Pain. 2004;8(3):201–10.

203. Cooney G, Dwan K, Mead G. Exercise for Depression. JAMA. 2014;311(23):2432–3.

204. Costa LDCM, Maher CG, McAuley JH, Hancock MJ, Smeets RJEM. Self-efficacy is more important than fear of movement in mediating the relationship between pain and disability in chronic low back pain. Eur J Pain. 2011;15(2):213–9.

205. Marshall PWM, Schabrun S, Knox MF. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related disability in people with chronic low back pain. Yang J, editor. PLoS One [Internet]. 2017 Jul 7 [cited 2018 Jul 9];12(7):e0180788. Available from: http://dx.plos.org/10.1371/journal.pone.0180788

206. Smith BE, Hendrick P, Bateman M, Holden S, Littlewood C, Smith TO, et al. Musculoskeletal pain and exercise - Challenging existing paradigms and introducing new. Br J Sports Med. 2019;53(14):907–12.

207. Mansell G, Hill JC, Main C, Vowles KE, van der Windt D. Exploring What Factors Mediate Treatment Effect: Example of the STarT Back Study High-Risk Intervention. J Pain [Internet]. 2016 Nov [cited 2017 Jun 30];17(11):1237–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27565304

208. Stevens ML, Boyle E, Hartvigsen J, Mansell G, Søgaard K, Jørgensen MB, et al. Mechanisms for reducing low back pain: a mediation analysis of a multifaceted intervention in workers in elderly care. Int Arch Occup Environ Health [Internet]. 2018;0(0):0. Available from: http://dx.doi.org/10.1007/s00420-018-1350-3

209. Froud R, Ellard D, Patel S, Eldridge S, Underwood M. Primary outcome measure use in back pain trials may need radical reassessment. BMC Musculoskelet Disord [Internet]. 2015 Dec 14 [cited 2017 Jun 19];16(1):88. Available from: http://download.springer.com/static/pdf/3/art%253A10.1186%252Fs12891-015-0534-1.pdf?originUrl=http%253A%252F%252Fbmcmusculoskeletdisord.biomedcentral.com%252Farticle%252F10.1186%252Fs12891-015-0534-1&token2=exp=1497880261~acl=%252Fstatic%252Fpdf%252F3%252

210. Harvey N, Holmes CA. Nominal group technique: An effective method for obtaining group consensus. Int J Nurs Pract. 2012;18(2):188–94.

211. Horton JN. Nominal group technique. Anaesthesia [Internet]. 1980 Aug 1 [cited 2018 Nov 1];35(8):811–4. Available from: http://doi.wiley.com/10.1111/j.1365-2044.1980.tb03924.x

212. Hill JC, Thomas E, Hill S, Foster NE, van der Windt DA. Development and Validation of the Keele Musculoskeletal Patient Reported Outcome Measure (MSK-PROM). Assassi S, editor. PLoS One [Internet]. 2015 Apr 30 [cited 2017 Jun 5];10(4):e0124557. Available from: http://dx.plos.org/10.1371/journal.pone.0124557

213. Humphrey-Murto S, Varpio L, Wood TJ, Gonsalves C, Ufholz LA, Mascioli K, et al. The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review. Acad Med. 2017;92(10):1491–8.

214. Scott EA, Black N. Appropriateness of cholecystectomy in the United Kingdom - A consensus panel approach. Gut. 1991;32(9):1066–70.

215. van Teijlingen E, Pitchforth E, Bishop C, Russell E. Delphi method and nominal group technique in family planning and reproductive health research. BMJ Sex Reprod Heal. 2006;32(4):249–52.

216. Lago PP, Beruvides MG, Jian JY, Canto AM, Sandoval A, Taraban R. Structuring group decision making in a web-based environment by using the nominal group technique. Comput Ind Eng. 2007;52(2):277–95.

217. Tillett W, Dures E, Hewlett S, Helliwell PS, FitzGerald O, Brooke M, et al. A multicenter nominal group study to rank outcomes important to patients, and their representation in existing composite outcome measures for psoriatic arthritis. J Rheumatol. 2017;44(10):1445–52.

218. Consolaro A, Ruperto N, Bazso A, Pistorio A, Magni-Manzoni S, Filocamo G, et al. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. Arthritis Care Res. 2009;61(5):658–66.

219. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite Outcomes in Cardiovascular Research: A Survey of Randomized Trials. Ann Intern Med [Internet]. 2008 Nov 4 [cited 2019 May 30];149(9):612. Available from: http://annals.org/article.aspx?doi=10.7326/0003-4819-149-9-200811040-00004

220. Saksena S, Skee A, Waldo AL, Freemantle N, Reynolds M, Rosenberg Y, et al. Cardiovascular outcomes in the AFFIRM trial: An assessment of individual antiarrhythmic drug therapies compared to rate control using propensity score matched analyses. J Am Coll Cardiol. 2011;58(19):1975–85.

221. Lubans D, Richards J, Hillman C, Faulkner G, Beauchamp M, Nilsson M, et al. Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms. Pediatrics [Internet]. 2016;138(3):e20161642–e20161642. Available from: http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2016-1642

222. Bombardier C, Evans CJ, Katz N, Mardekian J, Zlateva G, Simon LS. Further Qualification of a Therapeutic Responder Index for Patients with Chronic Low Back Pain. J Rheumatol [Internet]. 2011 Feb 1 [cited 2017 May 11];38(2):362–9. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/21041275

223. Coates LC, Lubrano E, Perrotta FM, Emery P, Conaghan PG, Helliwell PS. What Should Be the Primary Target of “Treat to Target” in Psoriatic Arthritis? J Rheumatol [Internet]. 2019 Jan [cited 2019 May 29];46(1):38–42. Available from: http://www.jrheum.org/lookup/doi/10.3899/jrheum.180267

224. Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. BMJ [Internet]. 2007;334(7597):756–7. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.39176.461227.80

225. Stenner R. Exercise prescription for patients with non-specific chronic low back pain : a qualitative exploration of physiotherapy practice. University of the West of England; 2014.

226. Mc Inerney SJ. Introducing the Biopsychosocial Model for good medicine and good doctors. BMJ. 2002;324:1537.

227. Sherman KJ, Wellman RD, Cook AJ, Cherkin DC, Ceballos RM. Mediators of yoga and stretching for chronic low back pain. Evid Based Complement Alternat Med [Internet]. 2013 Apr 17 [cited 2019 Feb 27];2013:130818. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23690832

228. Woby SR, Urmston M, Watson PJ. Self-efficacy mediates the relation between pain-related fear and outcome in chronic low back pain patients. Eur J Pain [Internet]. 2007 Oct [cited 2017 Jun 29];11(7):711–8. Available from: https://s3.amazonaws.com/objects.readcube.com/articles/downloaded/wiley/e83a0e32387173cc9eced5fb49a14a4c46cf05f3388a87e369eb4be19070659b.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIS5LBPCM5JPOCDGQ%2F20170629%2Fus-east-1%2Fs3%2Faws4_request&

229. Mansell G, Hill JC, Main CJ, Von Korff M, van der Windt D. Mediators of Treatment Effect in the Back In Action Trial. Clin J Pain [Internet]. 2017 Sep [cited 2019 Jul 2];33(9):811–9. Available from: http://insights.ovid.com/crossref?an=00002508-201709000-00006

230. De Silva MJ, Breuer E, Lee L, Asher L, Chowdhary N, Lund C, et al. Theory of Change:

a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. Trials [Internet]. 2014;15(1):267. Available from: http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-15-267

231. Lee H, Lamb SE. Advancing Physical Therapist Interventions by Investigating Causal Mechanisms. Phys Ther [Internet]. 2017 Dec 1 [cited 2018 Jun 13];97(12):1119–21. Available from: http://academic.oup.com/ptj/article/97/12/1119/4600554

232. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and prescription. Tenth. Riebe D, editor. Wolters Kluwer Health; 2017.

233. Littlecott HJ, Moore GF, Moore L, Murphy S. Psychosocial mediators of change in physical activity in the Welsh national exercise referral scheme: secondary analysis of a randomised controlled trial. Int J Behav Nutr Phys Act [Internet]. 2014 Dec 27 [cited 2018 Jul 4];11(1):109. Available from: http://ijbnpa.biomedcentral.com/articles/10.1186/s12966-014-0109-9

234. Wood L, Foster NE, Lewis M, Bishop A. Exercise interventions for persistent non-specific low back pain – does matching outcomes to treatment targets make a difference? A systematic review and meta-analysis. J Pain. 2020;in press.

235. Boivin A, Currie K, Gracia J, James M, Marshall C, Sakala C, et al. Patient and public involvement in clinical guidelines: international experiences and future perspectives. Qual Saf Heal Care. 2010;19(5):e22.

236. Hall AM, Kamper SJ, Emsley R, Maher CG. Does pain-catastrophising mediate the effect of tai chi on treatment outcomes for people with low back pain? Complement Ther Med [Internet]. 2016 Apr 1 [cited 2018 Jun 13];25:61–6. Available from: https://www.sciencedirect.com/science/article/pii/S0965229915300340?via%3Dihub

237. Vo TT, Superchi C, Boutron I, Vansteelandt S. The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. J Clin Epidemiol [Internet]. 2020;117:78–88. Available from: https://doi.org/10.1016/j.jclinepi.2019.10.001

238. Sherman KJ, Cherkin DC, Cook AJ, Hawkes RJ, Deyo RA, Wellman R, et al. Comparison of yoga versus stretching for chronic low back pain: Protocol for the Yoga Exercise Self-care (YES) trial. Trials. 2010;11:1–17.

239. Carpenter DM, Nelson BW. Low back strengthening for the prevention and treatment of low back pain. Med Sci Sports Exerc. 1999;31(1):18–24.

240. Radaelli R, Fleck SJ, Leite T, Leite RD, Pinto RS, Fernandes L, et al. Dose-response of 1,3, and 5 sets of resistance exercise on strength, local muscular endurance, and hypertrophy. J Strength Cond Res. 2015;29(5):1349–58.

241. Bendix AF, Bendix T, Vaegter K, Lund C, Frolund L, Holm L. Multidisciplinary intensive treatment for chronic low back pain: a randomized, prospective study. Cleve Clin J Med [Internet]. 1996;63(1):62–9. Available from: NS -

242. McIlveen B, Robertson VJ. A randomised controlled study of the outcome of hydrotherapy for subjects with low back or back and leg pain. Physiotherapy [Internet]. 1998;84(1):17–26. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=28031834 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:&id=doi:10.1016%2FS0031-9406%252805%252965898-1&issn=0031-9406&isbn=&volume=84&issue=1&spage=1

243. Jensen IB, Bergstrom G, Ljungquist T, Bodin L. A 3-year follow-up of a multidisciplinary rehabilitation programme for back and neck pain. Pain [Internet]. 2005;115(3):273–83. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=40725284 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:15911154&id=doi:10.1016%2Fj.pain.2005.03.005&issn=0304-3959&isbn=&volume=115&issue=3&spage=27

244. Jensen RK, Leboeuf-Yde C, Wedderkopp N, Sorensen JS, Manniche C. Rest versus exercise as treatment for patients with low back pain and Modic changes. a randomized controlled clinical trial. BMC Med [Internet]. 2012;10(1):22. Available from: http://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-10-22

245. Sherman K, Cherkin D, Wellman R, Cook A, Hawkes R, Delaney K, et al. A randomized trial comparing yoga, stretching, and a self-care book for chronic low back pain. Arch

Intern Med 2011 Dec [Internet]. 2011;12;171(22 PG-). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=amed&AN=0150447 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:ameddb&id=pmid:&id=doi:&issn=0003-9926&isbn=&volume=171&issue=22&spage=2019&pages=2019-26&date=2011&title=Archi

246. Sherman KJ, Cherkin DC, Erro J, Miglioretti DL, Deyo RA. Comparing yoga, exercise, and a self-care book for chronic low back pain: a randomized, controlled trial. Ann Intern Med [Internet]. 2005;143(12 PG-849–18):818–49. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106403857&site=ehost-live NS -

247. Smeets RJEM, Vlaeyen JWS, Hidding A, Kester ADM, Van Der Heijden GJMG, Van Geel ACM, et al. Active rehabilitation for chronic low back pain: Cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial ISRCTN22714229]. BMC Musculoskelet Disord [Internet]. 2006;7(PG-). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=43302332 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:16426449&id=doi:10.1186%2F1471-2474-7-5&issn=1471-2474&isbn=&volume=7&issue=1&spage=5&pages=&

248. Suni J, Rinne M, Natri A, Statistisian MP, Parkkari J, Alaranta H. Control of the Lumbar Neutral Zone Decreases Low Back Pain and Improves Self-Evaluated Work Ability. Spine (Phila Pa 1976) [Internet]. 2006;31(18):E611–20. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106161826&site=ehost-live NS -

249. Sokunbi O, Watt P, Moore A. A randomised controlled trial (RCT) on the effects of frequency of application of spinal stabilisation exercises on multifidus cross sectional area (MFCSA) in participants with chronic low back pain. Physiother Singapore [Internet]. 2008;11(2 PG-9–16):9–16. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed1

1&AN=352277947

http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:&id=doi:&issn=0219-1091&isbn=&volume=11&issue=2&spage=9&pages=9-16&date=2008&title=Physiothe

250. Van Der Roer N, Van Tulder M, Barendse J, Knol D, Van Mechelen W, De Vet H. Intensive group training protocol versus guideline physiotherapy for patients with chronic low back pain: A randomised controlled trial. Eur Spine J [Internet]. 2008;17(PG-):1193–200. Available from: NS -

251. Henchoz Y, de Goumoens P, So AKL, Paillex R. Functional multidisciplinary rehabilitation versus outpatient physiotherapy for non specific low back pain: randomized controlled trial. Swiss Med Wkly [Internet]. 2010;140(PG-). Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med6&AN=21181567 NS -

252. Hartvigsen J, Morsø L, Bendix T, Manniche C. Supervised and non-supervised Nordic walking in the treatment of chronic low back pain: a single blind randomized clinical trial. BMC Musculoskelet Disord [Internet]. 2010;11(1):30. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-11-30

253. Kamioka H, Okuizumi H, Okada S, Takahashi R, Handa S, Kitayuguchi J, et al. Effectiveness of intervention for low back pain in female caregivers in nursing homes: a pilot trial based on multicenter randomization. Environ Health Prev Med [Internet]. 2011;16(2):97–105. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=21432224 NS -

254. Murtezani A, Hundozi H, Orovcanec N, Sllamniku S, Osmani T. A comparison of high intensity aerobic exercise and passive modalities for the treatment of workers with chronic low back pain: A randomized, controlled trial. Eur J Phys Rehabil Med [Internet]. 2011;47(3):359–66. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=21602759 NS -

255. Miyamoto GC, Costa LOP, Galvanin T, Cabral CMN. Efficacy of the addition of modified

Pilates exercises to a minimal intervention in patients with chronic low back pain: a randomized controlled trial. Phys Ther [Internet]. 2013;93(3):310–20. Available from: NS -

256. Weifen W, Muheremu A, Chaohui C, MD LW, Lei S, Wenge L, et al. Effectiveness of Tai Chi Practice for Non-Specific Chronic Low Back Pain on Retired Athletes: A Randomized Controlled Study. J Musculoskelet Pain [Internet]. 2013;21(1):37–45. Available from: http://www.tandfonline.com/doi/full/10.3109/10582452.2013.763394

257. Aboagye E, Karlsson ML, Hagberg J, Jensen I. Cost-Effectiveness of Early Interventions for Non-Specific Low Back Pain: a Randomized Controlled Study Investigating Medical Yoga, Exercise Therapy and Self-Care Advice. J Rehabil Med (Stiftelsen Rehabiliteringsinformation) [Internet]. 2015;47(2 PG-167–173):167–73. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=107776610&site=ehost-live NS -

258. Akodu A, Akinbo RA, Odebiyi D. Effect of Stabilization Exercise on Pain and Disability in Patients with Chronic Low Back Pain. Indian J Physiother Occup Ther [Internet]. 2015;9(2):170–5. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=109813106&site=ehost-live NS -

259. Cruz-Díaz D, Martínez-Amat A, De La Torre-Cruz MJ, Casuso RA, De Guevara NML, Hita-Contreras F. Effects of a six-week Pilates intervention on balance and fear of falling in women aged over 65 with chronic low-back pain: A randomized controlled trial. Maturitas [Internet]. 2015;82(4):371–6. Available from: http://dx.doi.org/10.1016/j.maturitas.2015.07.022

260. Cruz-Diaz D, Martinez-Amat A, Osuna-Perez M, de la Torre-Cruz MJ, Hita-Contreras F. Short- and long-term effects of a six-week clinical Pilates program in addition to physical therapy on postmenopausal women with chronic low back pain: a randomized controlled trial with consumer summary]. Disabil Rehabil [Internet]. 2016;38(13):1300–8. Available from: NS -

261. Andersen LN, Juul-Kristensen B, Sörensen TL, Herborg LG, Roessler KK, Sögaard K. Longer term follow-up on effects of Tailored Physical Activity or Chronic Pain Self-

Management Programme on return-to-work: a randomized controlled trial. J Rehabil Med [Internet]. 2016;48(10 PG-887–892):887–92. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/158/CN-01263158/frame.html NS -

262. Chan AYP, Ford JJ, Surkitt LD, Richards MC, Slater SL, Davidson M, et al. Individualised functional restoration plus guideline-based advice vs advice alone for non-reducible discogenic low back pain: A randomised controlled trial. Physiother (United Kingdom) [Internet]. 2016;103(2):121–30. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=122947639&site=ehost-live NS -

263. Brämberg EB, Bergström G, Jensen I, Hagberg J, Kwak L, Bramberg EB, et al. Effects of yoga, strength training and advice on back pain: a randomized controlled trial. BMC Musculoskelet Disord [Internet]. 2017;18(1):132. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-017-1497-1

264. Cruz-Díaz D, Martínez-Amat A, Hita-Contreras F, Bergamin M, Gobbo S. Comparative effects of 12 weeks of equipment based and mat Pilates in patients with Chronic Low Back Pain on pain, function and transversus abdominis activation. A randomized controlled trial. Complement Ther Med [Internet]. 2017;33(February):72–7. Available from: http://dx.doi.org/10.1016/j.ctim.2017.06.004

265. Gelalis KG. Spinal mobilization versus conventional physiotherapy in the management of chronic low back pain due to spinal disk degeneration: a randomized controlled trial. J Man Manip Ther 2017;25(2)66-73. 2017;

266. Jayaseelan DH. Effectiveness of multipurpose health-worker-led exercise therapy on pain reduction among patients with chronic nonspecific low backache in primary health-care setting: a randomized control trial. J Fam Med Prim Care. 2019;8(1):199–202.

267. Shariat A, Cleland JA, Danaee M, Kargarfard M, Snagelaji B, Tamrin S, et al. Effects of stretching exercise training and ergonomic modifications on musculoskeletal discomforts of office workers: a randomized controlled trial. Brazilian J Phys Ther. 2018;22(2):144–53.

268. Suni JH, Kolu P, Tokola K, Raitanen J, Rinne M, Taulaniemi A, et al. Effectiveness and

cost-effectiveness of neuromuscular exercise and back care counseling in female healthcare workers with recurrent non-specific low back pain: a blinded four-arm randomized controlled trial. BMC Public Health. 2018;18(1):1376.

269. Kellett KM, Kellett DA, Nordholm LA. Effects of an Exercise Program on Sick Leave due to Back Pain. Phys Ther [Internet]. 1991;71(4 PG-283–291):283–91. Available from: NS -

270. Gundewall B, Liljeqvist M, Hansson T. Primary prevention of back symptoms and absence from work. A prospective randomized study among hospital employees. Spine (Phila Pa 1976) [Internet]. 1993;18(5):587–94. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/863/CN-00092863/frame.html NS -

271. Moffett JK, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Phillips H, Farrin A, et al. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs and preferences. BMJ [Internet]. 1999 [cited 2015 Jan 20];319:279–83. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=107227554&site=ehost-live NS -

272. Kuukkanen T, Mälkiä E. Effects of a three-month therapeutic exercise programme on flexibility in subjects with low back pain. Physiother Res Int [Internet]. 2000 Mar 1 [cited 2017 Aug 11];5(1):46–61. Available from: http://doi.wiley.com/10.1002/pri.183

273. Glomsrød B, Lønn JH, Soukup MG, Bø K, Larsen S. “Active back school”, prophylactic management for low back pain: three-year follow-up of a randomized, controlled trial. J Rehabil Med [Internet]. 2001;33(6):26–30. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=107040457&site=ehost-live NS -

274. Staal JB, Hlobil H, Twisk JW. WR, Smid T, Köke a. J. JA, van Mechelen W. Graded activity for low back pain in occupational health care: a randomized, controlled trial. Ann Intern Med [Internet]. 2004;140(140):77–84. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106731971&site=ehost-live NS -

275. Hlobil H, Staal JB, Twisk J, Köke A, Ariëns G, Smid T, et al. The effects of a graded

activity intervention for low back pain in occupational health on sick leave, functional status and pain: 12-month results of a randomized controlled trial. J Occup Rehabil [Internet]. 2005;15(4):569–80. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106340501&site=ehost-live

276. Jones GT, Johnson RE, Wiles NJ, Chaddock C, Potter RG, Roberts C, et al. Patient treatment preference influences outcome of low back pain: A randomised controlled trial of a community-based programme of exercise, education and cognitive behavioural therapy. Rheumatology [Internet]. 2006;45(PG-). Available from: NS -

277. NelsonWong E, Callaghan JP, Nelson-Wong E, Callaghan JP. Changes in muscle activation patterns and subjective low back pain ratings during prolonged standing in response to an exercise intervention. J Electromyogr Kinesiol [Internet]. 2010;20(6):1125–33. Available from: http://dx.doi.org/10.1016/j.jelekin.2010.07.007

278. Pedersen MT, Andersen CH, Zebis MK, Sjøgaard G, Andersen LL. Implementation of specific strength training among industrial laboratory technicians: long-term effects on back, neck and upper extremity pain. BMC Musculoskelet Disord [Internet]. 2013;14(1):287. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-14-287

279. Pedersen P, Nielsen CV, Jensen OK, Jensen C, Labriola M. Employment status five years after a randomised controlled trial comparing multidisciplinary and brief intervention in employees on sick leave due to low back pain. Scand J Public Health. 2018;46(3):383–8.

280. Kamada M, Kitayuguchi J, Abe T, Taguri M, Inoue S, Ishikawa Y, et al. Community-wide promotion of physical activity in middle-aged and older Japanese: a 3-year evaluation of a cluster randomized trial. Int J Behav Nutr Phys Act [Internet]. 2015;12(1):82. Available from: http://www.ijbnpa.org/content/12/1/82

281. Rasmussen CDN, Holtermann A, Bay H, Sogaard K, Birk Jorgensen M. A multifaceted workplace intervention for low back pain in nurses' aides: a pragmatic stepped wedge cluster randomised controlled trial. Pain [Internet]. 2015;156(9):1786–94. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&

AN=25993549 NS -

282. Irvine AB, Russell H, Manocchia M, Mino DE, Glassen TC, Morgan R, et al. Mobile-Web app to self-manage low back pain: Randomized controlled trial. J Med Internet Res [Internet]. 2015;17(1):1–21. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=psyh&AN=2016-31977-002&site=ehost-live NS -

283. Chaleat-Valayer E, Denis A, Abelin-Genevois K, Zelmar A, Siani-Trebern F, Touzet S, et al. Long-term effectiveness of an educational and physical intervention for preventing low-back pain recurrence: A randomized controlled trial. Scand J Work Environ Heal [Internet]. 2016;42(6):510–9. Available from: NS -

284. Rasmussen CDN, Holtermann A, Bay H, Søgaard K, Birk Jørgensen M. A multifaceted workplace intervention for low back pain in nurses' aides. Pain [Internet]. 2015;156(9):1786–94. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006396-201509000-00024

285. Suni JH, Rinne M, Tokola K, Mänttäri A, Vasankari T. Effectiveness of a standardised exercise programme for recurrent neck and low back pain: A multicentre, randomised, two-arm, parallel group trial across 34 fitness clubs in Finland. BMJ Open Sport Exerc Med. 2017;3(1):1–12.

286. Haufe S, Wiechmann K, Stein L, Kück M, Smith A, Meineke S, et al. Low-dose, non-supervised, health insurance initiated exercise for the treatment and prevention of chronic low back pain in employees. Results from a randomized controlled trial. PLoS One [Internet]. 2017 [cited 2017 Aug 17];12(6):e0178585. Available from: http://dx.plos.org/10.1371/journal.pone.0178585

287. Berglund E, Anderzén I, Andersén Å, Carlsson L, Gustavsson C, Wallman T, et al. Multidisciplinary Intervention and Acceptance and Commitment Therapy for Return-to-Work and Increased Employability among Patients with Mental Illness and/or Chronic Pain: A Randomized Controlled Trial. Int J Environ Res Public Health [Internet]. 2018 Oct 31 [cited 2019 Sep 25];15(11):2424. Available from: http://www.mdpi.com/1660-4601/15/11/2424

288. Wiklund T, Linton SJ, Alfoldi P, Gerdle B. Is sleep disturbance in patients with chronic pain affected by physical exercise or ACT-based stress management? - A randomized controlled study. BMC Musculoskelet Disord. 2018;19:111.

289. Elsenbruch S, Roderigo T, Enck P, Benson S. Can a Brief Relaxation Exercise Modulate Placebo or Nocebo Effects in a Visceral Pain Model? Front Psychiatry [Internet]. 2019 Mar 21 [cited 2019 Sep 25];10. Available from: https://www.frontiersin.org/article/10.3389/fpsyt.2019.00144/full

290. Johnsen TL, Eriksen HR, Baste V, Indahl A, Odeen M, Tveito TH. Effect of Reassuring Information About Musculoskeletal and Mental Health Complaints at the Workplace: A Cluster Randomized Trial of the atWork Intervention. J Occup Rehabil [Internet]. 2019 Jun 21 [cited 2019 Sep 25];29(2):274–85. Available from: http://link.springer.com/10.1007/s10926-018-9786-6

291. Vagnoni E, Biavati GR, Felisatti M, Pomidori L. Moderating healthcare costs through an assisted physical activity programme. Int J Health Plann Manage. 2018;33(4):1146–58.

292. Koes BW, Bouter LM, Van Mameren H, Essers AHM, Verstegen GMJR, Hofhuizen DM, et al. Third prize: A blinded randomized clinical trial of manual therapy and physiotherapy for chronic back and neck complaints: Physical outcome measures. J Manipulative Physiol Ther [Internet]. 1992;15(1):16–23. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed5&AN=22080741 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:1531487&id=doi:&issn=0161-4754&isbn=&volume=15&issue=1&spage=16&pages=16-23&date=1992&title=Jo

293. Skouen JS, Grasdal AL, Haldorsen EMH, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine (Phila Pa 1976) [Internet]. 2002;27(9):901–10. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106974027&site=ehost-live NS -

294. Vollenbroek-Hutten MM, Hermens HJ, Wever D, Gorter M, Rinket J, Ijzerman MJ.

Differences in outcome of a multidisciplinary treatment between subgroups of chronic low back pain patients defined using two multiaxial assessment instruments: the multidimensional pain inventory and lumbar dynamometry with consumer summary]. Clin Rehabil [Internet]. 2004;18(PG-):566–79. Available from: http://journals.sagepub.com/doi/10.1191/0269215504cr772oa

295. Frost H, Lamb SE, Doll HA, Carver PT, Stewart-Brown S. Randomised controlled trial of physiotherapy compared with advice for low back pain. BMJ [Internet]. 2004;329(7468):708–11. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106600856&site=ehost-live

296. Lindell O, Johansson SE, Strender LE, Lindell O, Johansson S-E, Strender L-E. Subacute and chronic, non-specific back and neck pain: cognitive-behavioural rehabilitation versus primary care. A randomized controlled trial. BMC Musculoskelet Disord [Internet]. 2008;9:172. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=105524153&site=ehost-live

297. Bültmann U, Sherson D, Olsen J, Hansen CL, Lund T, Kilsgaard J. Coordinated and tailored work rehabilitation: a randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. J Occup Rehabil [Internet]. 2009;19(1):81–93. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=105458419&site=ehost-live NS -

298. Ernst MJ, Crawford RJ, Schelldorfer S, Rausch-Osthoff A-K, Barbero M, Kool J, et al. Extension and flexion in the upper cervical spine in neck pain patients. Man Ther [Internet]. 2014 Dec 31 [cited 2015 May 11]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/25578386

299. Marchand GH, Myhre K, Leivseth G, Sandvik L, Lau B, Bautz-Holter E, et al. Change in pain, disability and influence of fear-avoidance in a work-focused intervention on neck and back pain: a randomized controlled trial. BMC Musculoskelet Disord [Internet]. 2015;16(1):94. Available from:

http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-015-0553-y

300. Kogure A, Kotani K, Katada S, Takagi H, Kamikozuru M, Isaji T, et al. A Randomized, Single-Blind, Placebo-Controlled Study on the Efficacy of the Arthrokinematic Approach-Hakata Method in Patients with Chronic Nonspecific Low Back Pain. PLoS One. 2015;10(12).

301. Tavafian SS, Jamshidi A, Shay B. Treatment of low back pain: First extended follow up of an original trial (NCT00600197) comparing a multidisciplinary group-based rehabilitation program with oral drug treatment alone up to 24 months. Int J Rheum Dis. 2017;20:1902–9.

302. Hahne AJ, Ford JJ, Surkitt LD, Richards MC, Chan AYP, Slater SL, et al. Individualized physical therapy is cost-effective compared with guideline-based advice for people with low back disorders. Spine (Phila Pa 1976). 2017;42(3):E169–76.

303. Malfliet A, Kregel J, Meeus M, Roussel N, Danneels L, Cagnie B, et al. Blended learning pain neuroscience education for people with chronic spinal pain: a randomized-controlled multi-centre trial. Phys Ther. 2017;Awaiting p:1–35.

304. Cherkin D, Balderson B, Wellman R, Hsu C, Sherman KJ, Evers SC, et al. Effect of Low Back Pain Risk-Stratification Strategy on Patient Outcomes and Care Processes: the MATCH Randomized Trial in Primary Care. J Gen Intern Med [Internet]. 2018;33(8):1–13. Available from: http://link.springer.com/10.1007/s11606-018-4468-9

305. Hasan S, Widyahening IS, Hirawan S. The Effectiveness of William's Flexion Movement Towards the Decrease in the Prevalence of Low Back Pain in Informal Shoe Production Workshops. Adv Sci Lett [Internet]. 2018 Sep 1 [cited 2019 Sep 25];24(9):6945–9. Available from: http://www.ingentaconnect.com/content/10.1166/asl.2018.12892

306. Carrasco-Martínez F, Ibáñez-Vera AJ, Martínez-Amat A, Hita-Contreras F, Lomas-Vega R. Short-term effectiveness of the flexion-distraction technique in comparison with high-velocity vertebral manipulation in patients suffering from low-back pain. Complement Ther Med [Internet]. 2019 Jun [cited 2019 Sep 25];44:61–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0965229918310136

307. Li H, Zhang H, Liu S, Wang Y, Gai D, Lu Q, et al. Rehabilitation effect of exercise with soft tissue manipulation in patients with lumbar muscle strain. Niger J Clin Pract.

2017;20(5):629–33.

308. Waseem M, Karimi H, Gilani SA, Hassan D. Treatment of disability associated with chronic non-specific low back pain using core stabilization exercises in Pakistani population. J Back Musculoskelet Rehabil. 2019;32(1):149–54.

309. Callaghan MJ. Evaluation of a back rehabilitation group for chronic low back pain in an out-patient setting. Physiotherapy [Internet]. 1994;80(10 PG-677–681):677–81. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed6&AN=24316509 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:&id=doi:&issn=0031-9406&isbn=&volume=80&issue=10&spage=677&pages=677-681&date=1994&title=Physi

310. Bendix AF, Bendix T, Labriola M, Boekgaard P. Functional restoration for chronic low back pain: two-year follow-up of two randomized clinical trials. Spine (03622436) [Internet]. 1998;23(6 PG-717–725):717–25. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=107262520&site=ehost-live NS -

311. Bendix AF, Bendix T, Haestrup C, Busch E. A prospective, randomized 5-year follow-up study of functional restoration in chronic low back pain patients. Eur Spine J [Internet]. 1998;7(2 PG-111–119):111–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=28223596 http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:9629934&id=doi:10.1007%2Fs005860050040&issn=0940-6719&isbn=&volume=7&issue=2&spage=111&pages=1

312. Carr JL, Klaber Moffett JA, Howarth E, Richmond SJ, Torgerson DJ, Jackson DA, et al. A randomized trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. Disabil Rehabil [Internet]. 2005;27(16):929–37. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed1

0&AN=41052438
http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:16096246&id=doi:10.1080%2F09638280500030639&issn=0963-8288&isbn=&volume=27&issue=16&spage=929

313. Critchley DJ, Ratcliffe J, Noonan S, Jones RH, Hurley M V. Effectiveness and cost-effectiveness of three types of physiotherapy used to reduce chronic low back pain disability: a pragmatic randomized trial with economic evaluation. Spine (Phila Pa 1976) [Internet]. 2007;32(14):1474–81. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00007632-200706150-00003

314. Frih ZBS, Fendri Y, Jellad A, Boudoukhane S, Rejeb N. Efficacy and treatment compliance of a home-based rehabilitation programme for chronic low back pain: A randomized, controlled study. Ann Phys Rehabil Med [Internet]. 2009;52(6):485–96. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed12&AN=50529126
http://nhs4315962.on.worldcat.org/atoztitles/link?sid=OVID:embase&id=pmid:19473905&id=doi:10.1016%2Fj.rehab.2009.04.002&issn=1877-0657&isbn=&volume=52&issue=6&spage=48

315. Henchoz Y, de Goumoens P, Norberg M, Paillex R, So AK. Role of physical exercise in low back pain rehabilitation: a randomized controlled trial of a three-month exercise program in patients who have completed multidisciplinary rehabilitation. Spine (Phila Pa 1976) [Internet]. 2010;35(12):1192–9. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20098350%5Cnhttp://graphics.tx.ovid.com/ovftpdfs/FPDDNCOBMECJGJ00/fs046/ovft/live/gv025/00007632/00007632-201005200-00003.pdf

316. Kumar S, Sharma VP, Shukla R, Dev R. Comparative efficacy of two multimodal treatments on male and female sub-groups with low back pain (part II). J Back Musculoskelet Rehabil [Internet]. 2010;23(1):1–9. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&A

N=105152297&site=ehost-live NS -

317. Ben-Ami N, Chodick G, Mirovsky Y, Pincus T, Shapiro Y. Increasing Recreational Physical Activity in Patients With Chronic Low Back Pain: A Pragmatic Controlled Clinical Trial. J Orthop Sport Phys Ther [Internet]. 2017;47(2):57–66. Available from: http://www.jospt.org/doi/10.2519/jospt.2017.7057

318. Thanawat T, Nualnetr N. Effects of an intervention based on the Transtheoretical Model on back muscle endurance, physical function and pain in rice farmers with chronic low back pain. J Back Musculoskelet Rehabil [Internet]. 2017;30(4):847–56. Available from: NS -

319. Mbada C, Olaoye M, Ayanniyi O, Johnson O, Odole A, Dada O. Comparative Efficacy of Clinic-Based and Telerehabilitation Application of Mckenzie Therapy in Low-Back Pain. Arch Phys Med Rehabil [Internet]. 2017 Oct [cited 2018 Aug 10];98(10):e46–7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003999317306925

320. Canaway A, Pincus T, Underwood M, Shapiro Y, Chodick G, Ben-Ami N. Is an enhanced behaviour change intervention cost-effective compared with physiotherapy for patients with chronic low back pain? Results from a multicentre trial in Israel. BMJ Open [Internet]. 2018 Apr 10 [cited 2019 Sep 25];8(4):e019928. Available from: http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-019928

321. Müller G, Pfinder M, Clement M, Kaiserauer A, Deis G, Waber T, et al. Therapeutic and economic effects of multimodal back exercise: A controlled multicentre study. J Rehabil Med [Internet]. 2019 [cited 2019 Sep 25];51(1):61–70. Available from: https://www.medicaljournals.se/jrm/content/abstract/10.2340/16501977-2497

322. Schulz C, Evans R, Maiers M, Schulz K, Leininger B, Bronfort G. Spinal manipulative therapy and exercise for older adults with chronic low back pain: a randomized clinical trial. Chiropr Man Therap [Internet]. 2019 Dec 15 [cited 2019 Sep 25];27(1):21. Available from: https://chiromt.biomedcentral.com/articles/10.1186/s12998-019-0243-1

323. Hurley DA, Tully MA, Lonsdale C, Boreham CAGG, van Mechelen W, Daly L, et al. Supervised walking in comparison with fitness training for chronic back pain in physiotherapy: results of the SWIFT single-blinded randomized controlled trial (ISRCTN17592092). Pain [Internet]. 2015;156(1):131–47. Available from: NS -

324. Alaranta H, Rytokoski U, Rissanen A, Talo S, Ronnemaa T, Puukka P, Karppi SL, Videman T, Kallio V SlP. Intensive physical and psychosocial training program for patients with chronic low back pain. A controlled clinical trial. Spine (Phila Pa 1976) [Internet]. 1994 [cited 2018 Aug 10];19(12):1339–49. Available from: https://ovidsp.tx.ovid.com/sp-3.31.1b/ovidweb.cgi?QS2=434f4e1a73d37e8c27e134f44f8715f17e8b1a06bc592dd5a14c83e2b537f0b3b73b6df06570b93d4842e9e9e71dacc69c6ab7c9f6a6808fb27b423d90e26dc0ed8a3f8256c24351d89114ff0234c6fb7b4d5ee040fe0a7fab926a1ae9c0cdb88282e5701

325. Guarino P, Peduzzi P, Donta ST, Engel CC, Clauw DJ, Williams DA, et al. A multicenter two by two factorial trial of cognitive behavioral therapy and aerobic exercise far gulf war veterans' illnesses: Design of a veterans affairs cooperative study (CSP #470). Control Clin Trials [Internet]. 2001;22(3):310–32. Available from: NS -

326. Norwegian RCT of 64 patients with chronic back pain. Bone Jt [Internet]. 2003;9(11):127–8. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=106753806&site=ehost-live NS -

327. Klaber Moffett JA, Carr J, Howarth E. High Fear-Avoiders of Physical Activity Benefit From an Exercise Program for Patients With Back Pain. Spine (Phila Pa 1976) [Internet]. 2004;29(11):1167–72. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00007632-200406010-00002

328. Watson P. No title. Talkback Mag [Internet]. 2008;(PG-24):24. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=105788674&site=ehost-live NS -

329. Gladkowski CA, Medley CL, Nelson HM, Price AT, Harvey M. Opioids Versus Physical Therapy for Management of Chronic Back Pain. J Nurse Pract [Internet]. 2014;10(8):552–9. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid,shib&db=rzh&AN=107828642&site=ehost-live

330. Ahmad AD. The efficacy of physiotherapy combination technique on pain and functional independence of people with lumbar disc herniation. Indian J Public Heal Res Dev. 2018;9(8):1524–9.

331. Saper RB, Sherman KJ, Delitto A, Herman PM, Stevans J, Paris R, et al. Yoga vs. physical therapy vs. education for chronic low back pain in predominantly minority populations: study protocol for a randomized controlled trial. BMC. 2014;15(67):1–41.

332. Leeuw M, Goossens M, van Breukelen G, de Jong J, Heuts P, Smeets R, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. Pain. 2008;138(1):192–207.

333. Seymour RB, Hughes SL, Campbell RT, Huber GM, Desai P. Comparison of Two Methods of Conducting Fit and Strong! Arthritis Rheum [Internet]. 2009 [cited 2018 Jun 13];61(7):876–84. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2710408/pdf/nihms117377.pdf

334. Smeets RJEM, Vlaeyen JW, Hidding A, Kester AD, van der Heijden GJ, van Geel AC, et al. Active rehabilitation for chronic low back pain: Cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial [ISRCTN22714229]. BMC Musculoskelet Disord [Internet]. 2006;7(1):5. Available from: http://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-7-5

335. Lethem J, Slade P, Troup J, Bentley G. Outline of a Fear-Avoidance Model of exaggerated pain perception. Behav Res Ther. 1983;21(4):401–8.

9 Appendices

a. Systematic Review Search Terms

MEDLINE	
Search terms for persistent NSLBP	
1	Exp Back/
2	Spine/ or Coccyx/ or Intervertebral Disc/ or Lumbar Vertebrae/ or Intervertebral disc/ or Sacrum/
3	(back or spine or spinal or lumb$ or sacr$ or coccy$).ti,ab,kw.
4	or/1-3
5	pain/ or chronic pain/ or pain, intractable/ or pain, referred/
6	4 and 5
7	(pain or painful).ti,ab,kw.
8	1 or 2
9	7 and 8
10	6 or 9
11	exp Back Pain/
12	(backache$ or (back adj3 ache$)).ti,ab,kw.
13	((back or spine or spinal or lumb$ or sacr$ or coccy$) adj3 pain$).ti,ab,kw.
14	lumbago.ti,ab,kw.
15	dorsalgia.ti,ab,kw.
16	coccydynia.ti,ab,kw.
17	(LBP or cLBP).ti,ab,kw.
18	or/10-17
Search terms for Exercise	
19	exp Exercise/
20	exp Rehabilitation/
21	Osteopathic Physicians/
22	Chiropractic/
23	(strength$ or isometric$ or isotonic$ or isokinetic$).ti,ab,kw.
24	(resistance adj3 train$).ti,ab,kw.
25	exercise$.ti,ab,kw.
26	((water$ or aqua$) adj3 (therap$ or treatment$)).ti,ab,kw.
27	(hydrotherap$ or aquatherap$).ti,ab,kw.
28	physiotherap$.ti,ab,kw.
29	osteopath$.ti,ab,kw.
30	chiropract$.ti,ab,kw.
31	physical therap$.ti,kw,ab.
32	rehabilitat$.ti,ab,kw.
33	19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
RCT filter – Cochrane sensitivity & precision max filter (2008 revision)	
34	randomized controlled trial.pt.
35	controlled clinical trial.pt.
36	randomi#ed.ab.
37	Placebo.ab.
38	clinical trials as topic/
39	randomly.ab.
40	trial.ti.
41	34 or 35 or 36 or 37 or 38 or 39 or 40 or 41
42	exp animals/ not humans/
43	41 not 42
44	18 and 33 and 43

EMBASE	
Search terms for persistent NSLBP	
1	exp back/
2	spine/ or coccyx/ or intervertebral disk/ or lumbar vertebra/ or sacrum/
3	(back or spine or spinal or lumb$ or sacr$ or coccy$).ti,ab,kw.
4	or/1-3
5	pain/ or referred pain/ or chronic pain/ or discogenic pain/ or intractable pain/
6	4 and 5
7	(pain or painful).ti,ab,kw.
8	1 or 2
9	7 and 8
10	6 or 9
11	exp backache/
12	exp low back pain/
13	(backache$ or (back adj3 ache$)).ti,ab,kw.
14	((back or spine or spinal or lumb$ or sacr$ or coccy$) adj3 pain$).ti,ab,kw.
15	lumbago.ti,ab,kw.
16	dorsalgia.ti,ab,kw.
17	coccydynia.ti,ab,kw.
18	(LBP or cLBP).ti,ab,kw.
19	or/10-18
Search terms for Exercise	
20	exp physiotherapy/
21	exp exercise/
22	osteopathic physician/
23	chiropractic/
24	(strength$ or isometric$ or isotonic$ or isokinetic$).ti,ab,kw.
25	(resistance adj3 train$).ti,ab,kw.
26	exercise$.ti,ab,kw.
27	((water$ or aqua$) adj3 (therap$ or treatment$)).ti,ab,kw.
28	(hydrotherap$ or aquatherap$).ti,ab,kw.
29	physiotherap$.ti,ab,kw.
30	osteopath$.ti,ab,kw.
31	chiropract$.ti,ab,kw.
32	physical therap$.ti,ab,kw.
33	rehabilitat$.ti,ab,kw.
34	or/20-33
RCT filter	
35	random$.tw.
36	factorial$.tw.
37	crossover$.tw.
38	cross-over$.tw.
39	placebo$.tw.
40	(doubl$ adj blind$).tw.
41	(singl$ adj blind$).tw.
42	assign$.tw.
43	allocat$.tw.
44	volunteer$.tw.
45	crossover procedure/
46	double blind procedure/
47	randomized controlled trial/
48	single blind procedure/
49	or/35-48
50	animal/ not human/
51	49 not 50
52	19 and 34 and 51

53	limit 52 to embase
AMED	
Search terms for persistent NSLBP	
1	exp Back/ 709
2	Spine/ or Intervertebral disk/ or Lumbar vertebrae/ or Sacrum/
3	(back or spine or spinal or lumb$ or sacr$ or coccy$).ti,ab.
4	or/1-3
5	Pain/
6	4 and 5
7	(pain or painful).ti,ab.
8	1 or 2
9	7 and 8
10	6 or 9
11	exp Low back pain/ or exp Backache/
12	((back adj3 ache$) or backache$).ti,ab.
13	((back or spine or spinal or lumb$ or sacr$ or coccy$) adj3 pain$).ti,ab.
14	lumbago.ti,ab.
15	dorsalgia.ti,ab.
16	coccydynia.ti,ab.
17	(LBP or cLBP).ti,ab.
18	or/10-17
	Search terms for Exercise
19	exp physical therapy modalities/
20	exp Exercise/
21	exp Rehabilitation/
22	Chiropractic/ or Osteopathy/
23	(strength$ or isometric$ or isotonic$ or isokinetic$).ti,ab.
24	(resistance adj3 train$).ti,ab.
25	exercise$.ti,ab.
26	((water$ or aqua$) adj3 (therap$ or treatment$)).ti,ab.
27	(hydrotherap$ or aquatherap$).ti,ab.
28	physiotherap$.ti,ab.
29	osteopath$.ti,ab.
30	chiropract$.ti,ab.
31	physical therap$.ti,ab.
32	rehabilitat$.ti,ab.
33	or/19-33
	RCT filter
34	exp clinical trials/
35	randomized controlled trial.pt.
36	controlled clinical trial.pt.
37	randomi#ed.ab.
38	placebo.ab.
39	randomly.ab.
40	trial.ti.
41	or/34-40
42	exp Animals/
43	exp humans/
44	42 not 43
45	41 not 44
46	18 and 33 and 45
Web of Science Core Collection	
Search terms for persistent NSLBP	
1	TS=((back OR spine OR spinal OR lumb* OR sacr* OR coccy*) NEAR/3 pain*)
2	TS= (backache* OR (back NEAR/3 ache*))
3	TS= (lumbago OR dorsalgia OR coccydynia)

4	TS=(LBP or CLBP)
5	#4 OR #3 OR #2 OR #1
Search terms for Exercise	
6	TS=exercise*
Search terms for RCTs	
7	TS=random*
8	TS=placebo*
9	TS=(clinic* NEAR/3 trial*)
10	#9 OR #8 OR #7
11	#10 AND #6 AND #5
PEDRO	
	Therapy: strength training
	Problem: Pain
	Body part: Lumbar spine, sacro-iliac joint or pelvis
	Subdiscipline: musculoskeletal
	Topic: Chronic pain
	Method: Clinical Trial
	Minimum score 6/10
	Therapy: hydrotherapy, balneotherapy
	Problem: Pain
	Body part: Lumbar spine, sacro-iliac joint or pelvis
	Subdiscipline: musculoskeletal
	Topic: Chronic pain
	Method: Clinical Trial
	Minimum score 6/10
	Therapy: fitness training
	Problem: Pain
	Body part: Lumbar spine, sacro-iliac joint or pelvis
	Subdiscipline: musculoskeletal
	Topic: Chronic pain
	Method: Clinical Trial
	Minimum score 6/10
	When searching Match all search terms
COCHRANE SEARCH	
Search terms for persistent NSLBP	
1	MeSH descriptor: [Back] explode all trees 640
2	MeSH descriptor: [Spine] this term only
3	MeSH descriptor: [Coccyx] this term only
4	MeSH descriptor: [Intervertebral Disc] this term only
5	MeSH descriptor: [Lumbar Vertebrae] this term only
6	MeSH descriptor: [Sacrum] this term only
7	(back or spine or spinal or lumb* or sacr* or coccy*):ti,ab,kw
8	#1 or #2 or #3 or #4 or #5 or #6 or #7
9	MeSH descriptor: [Pain] this term only
10	MeSH descriptor: [Chronic Pain] this term only
11	MeSH descriptor: [Pain, Intractable] this term only
12	MeSH descriptor: [Pain, Referred] this term only
13	#9 or #10 or #11 or #12
14	#8 and #13
15	(pain or painful):ti,ab,kw
16	#1 or #2 or #3 or #4 or #5 or #6
17	#15 and #16
18	#14 or #17

Search terms for Exercise	
19	MeSH descriptor: [Physical Therapy Modalities] explode all trees
20	MeSH descriptor: [Exercise] explode all trees
21	MeSH descriptor: [Rehabilitation] explode all trees
22	MeSH descriptor: [Osteopathic Physicians] this term only
23	MeSH descriptor: [Chiropractic] this term only
24	(strength* or isometric* or isotonic* or isokinetic*):ti,ab,kw
25	(resistance near/3 train*):ti,ab,kw
26	exercise*:ti,ab,kw
27	((water* or aqua*) near/3 (therap* or treatment*)):ti,ab,kw
28	(hydrotherap* or aquatherap*):ti,ab,kw
29	physiotherap*:ti,ab,kw
30	osteopath*:ti,ab,kw
31	chiropract*:ti,ab,kw
32	physical next therap*:ti,ab,kw
33	rehabilitat*:ti,ab,kw
34	#19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33
35	#18 and #34
Search terms for RCTs	
36	Trials only
CINAHL PLUS (EBSCO)	
Search terms for persistent NSLBP	
S1	(MH "Back")
S2	TI ((back or spine or spinal or lumb* or sacr* or coccy*)) OR AB ((back or spine or spinal or lumb* or sacr* or coccy*))
S3	S1 OR S2
S4	(MH "Pain+") OR (MH "Chronic Pain")
S5	S3 AND S4
S6	TI ((pain or painful)) OR AB ((pain or painful))
S7	S1 AND S6
S8	S5 OR S7
S9	(MH "Back Pain+") OR (MH "Low Back Pain")
S10	TI ((backache or (back n3 ache))) OR AB ((backache or (back n3 ache)))
S11	TI lumbago OR AB lumbago
S12	TI dorsalgia OR AB dorsalgia
S13	TI coccydynia OR AB coccydynia
S14	TI ((LBP or CLBP)) OR AB ((LBP or CLBP))
S15	S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14
Search terms for Exercise	
S16	(MH "Exercise+") (80,241)
S17	(MH "Psychotherapy+") OR (MH "Cognitive Therapy+") (137,615)
S18	(MH "Rehabilitation+") (214,392)
S19	(MH "Osteopathic Medicine") (331)
S20	(MH "Physical Therapy+") (104,221)
S21	TI ((strength* OR isometric* OR isotonic* OR isokinetic*)) OR AB ((strength* OR isometric* OR isotonic* OR isokinetic*)) (71,881)
S22	TI ("resistance train*") OR AB ("resistance train*") (3,401)
S23	TI exercise* OR AB exercise* (79,668)
S24	TI ((hydrotherap* OR aquatherap*) OR ((aqua* OR water*) n3 (therap* OR treatment*))) OR AB ((hydrotherap* OR aquatherap*) OR ((aqua* OR water*) n3 (therap* OR treatment*))) (1,679)
S25	TI physiotherap* OR AB physiotherap* (14,669)
S26	TI osteopath* OR AB osteopath* (2,374)
S27	TI chiropract* OR AB chiropract* (11,473)
S28	(MH "Chiropractors") (3,337)

S29	(MH "Osteopaths") 317)
S30	TI ("physical therap*") OR AB ("physical therap*") (14,350)
S31	TI rehabilitat* OR AB rehabilitat* (66,748)
S32	S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31
Search terms for RCTs	
S33	TX randomized OR (MH "Treatment Outcomes") OR PT (clinical trial)
S34	S15 and S32 and S33
Psychinfo search	
Search terms for persistent NSLBP	
S1	DE "Spinal Column"
S2	TI ((back or spine or spinal or lumb* or sacr* or coccy*)) OR AB ((back or spine or spinal or lumb* or sacr* or coccy*)) OR KW ((back or spine or spinal or lumb* or sacr* or coccy*))
S3	S1 OR S2
S4	DE "Pain" OR DE "Chronic Pain" OR DE "Myofascial Pain"
S5	S3 AND S4
S6	TI ((pain or painful)) OR AB ((pain or painful)) OR KW ((pain or painful))
S7	S1 AND S6
S8	S5 OR S7
S9	DE "Back Pain"
S10	TI ((backache* OR (back n3 ache*))) OR AB ((backache* OR (back n3 ache*))) OR KW ((backache* OR (back n3 ache*)))
S11	TI (((back or spine or spinal or lumb* or sacr* or coccy*) n3 pain*)) OR AB (((back or spine or spinal or lumb* or sacr* or coccy*) n3 pain*)) OR KW (((back or spine or spinal or lumb* or sacr* or coccy*) n3 pain*))
S12	TI lumbago OR AB lumbago OR KW lumbago
S13	TI dorsalgia OR AB dorsalgia OR KW dorsalgia
S14	TI coccydynia OR AB coccydynia OR KW coccydynia
S15	TI ((LBP or cLBP)) OR AB ((LBP or cLBP)) OR KW ((LBP or cLBP))
S16	S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15
Search terms for Exercise	
S17	DE "Exercise" OR DE "Aerobic Exercise" OR DE "Weightlifting" OR DE "Yoga" (27,916)
S18	DE "Rehabilitation" OR DE "Physical Therapy" OR DE "Psychosocial Rehabilitation" (24,321)
S19	DE "Osteopathic Medicine" (131)
S20	DE "Physical Therapy" (2,433)
S21	TI ((strength* or isometric* or isotonic* or isokinetic*)) OR AB ((strength* or isometric* or isotonic* or isokinetic*)) OR KW ((strength* or isometric* or isotonic* or isokinetic*)) (110,138)
S22	TI ("resistance train*") OR AB ("resistance train*") OR KW ("resistance train*") (564)
S23	TI exercise* OR AB exercise* OR KW exercise* (57,728)
S24	TI ((hydrotherap* or aquatherap*) OR ((water* or aqua*) n3 (treatment* or therap*))) OR AB ((hydrotherap* or aquatherap*) OR ((water* or aqua*) n3 (treatment* or therap*)))OR KW ((hydrotherap* or aquatherap*) OR ((water* or aqua*) n3 (treatment* or therap*))) (694)
S25	TI physiotherap* OR AB physiotherap* OR KW physiotherap* (2,682)
S26	TI osteopath* OR AB osteopath* OR KW osteopath* (289)
S27	TI chiropract* OR AB chiropract* OR KW chiropract* (387)
S28	TI ("physical therap*") OR AB ("physical therap*") OR KW ("physical therap*") (3,124)
S29	TI rehabilitat* OR AB rehabilitat* OR KW rehabilitat* (54,954)
S30	S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29
Search terms for RCTs	
S31	TI double-blind OR AB double-blind

S32	TI random* w1 assigned OR AB random* w1 assigned
S33	TI control OR AB control
S34	S31 OR S32 OR S33
S35	S16 AND S30 AND S34

b. Excluded Studies from the Systematic Review with Reasons

Reason for Exclusion	**References**
Insufficient sample size (n=29)	65,233–260
Incorrect population (n=23)	259,260,269–278,261,279–281,262–268
Incorrect Intervention (n=18)	275,295,304–311,296–303
Active Comparator (n=15)	305,306,315–319,307–314
Incorrect trial design (n=6)	289–294
Unavailable (n=1)	312

c. Extracted Treatment Targets and Outcomes of Included Trials

Author	Treatment Targets	Primary Outcome	Matched	Specified/ Inferred
Albadejo et al. (116)	Not mentioned	Physical function (RMDQ)	No	None
Bronfort et al. (117)	Exercises designed to increase *trunk muscle endurance and trunk stability;* Manipulation to improve function.	Patient rated pain intensity	No	Specified
Cambron et al. (118)	Aim of the programme was to *strengthen* the muscles surrounding the sine and *increase flexibility*.	Pain intensity, Physical function, Health-related quality of life.	No	Specified
Chen et al. (120)	*Reduce LBP and improve physical capability*. Effectiveness of SEP on *exercise self-efficacy*	*Pain*	Yes	Inferred
Chown et al. (121)	Not mentioned	Physical function	No	None
Costa et al. (122)	Exercises designed to *improve function of specific muscles of the LB* and the *control of posture and movement*	Pain intensity	No	Specified
Diaz Arribas et al. (123)	*Postural or functional* overload causes *biomechanical alteration*	Pain	No	Specified
Ferreira et al. (124)	General exercise: Reverse deconditioning of the *fear of movement* associated with pain Motor control exercise: *Retrain optimal control* of spinal motion	Physical function and global perceived effect of treatment	No	Specified
Garcia et al. (125)	“identified by *relief of pain*”; “to *stretch shortened tissues*”	*Pain* and disability	Yes	Specified
Goldby et al. (126)	The exercise aimed to rehabilitate the *neural control and active subsystems of the lumbar spine's stabilising system*. Selective retraining of the TrA, Multifidus, pelvic floor, diaphragm, and inhibit global muscle substitution mechanisms.	Pain	No	Specified
Groessl et al. (127)	*Increased strength and flexibility* and indirectly through the effects of breathing and meditation techniques that promote *stress reduction* and increased parasympathetic tone which in turn modulates *pain tolerance*.	Physical function (RMDQ)	No	Specified
Hall et al. (129)	*Balance, strengthening, stretching, body awareness, general health and well-being*	Bothersome-ness of pain	No	Specified
Hansen et al. (128)	Unclear but mentions "tense back muscles being stretched through exercise" and "body building", "hyperextending back exercises", so muscle tension, muscle strength, load on	Pain	No	Inferred

	and movement of the spine, muscle coordination and control and static trunk muscle endurance? But the targets are not clear.			
Harris et al. (130)	The introduction mentions several targets for brief interventions (biopsychosocial factors, believed to lower fear-avoidance and increase belief in recovery). Increase physical activity, *decrease pain improve function*; "the goal was to address *fear-avoidance* and movement phobia, and *help to re-establish normal movement patterns."* They later mention muscle hypertrophy and higher aerobic capacity.	Increased work participation	No	Specified
Hildebrandt et al. (131)	*Posture and movement patterns, recovery*	*Self-recovery*	No	Inferred
Jans et al. (132)	None mentioned	Self-recovery	No	None
Jaromi et al. (133)	Improvement in the execution of proper patient lifting techniques, reduce low back pain	Lifting techniques and Pain intensity	Yes	Specified
Johnson et al. (134)	Changes in psychosocial factors. The paper states 'chronic LBP and associated disability is multifactorial in aetiology and best understood with a biopsychosocial model. Studies have shown the main determinants of disability in LBP are p*sychosocial* and changes in psychosocial factors appear to be key in reduction of reported disability in people undergoing physiotherapy exercise programmes'.	Physical Function	No	Inferred
Maul et al. (135)	Deconditioning syndrome –*improve muscular stabilisation and functional capacity*; *increasing level of activity and self-confidence*; modifying *perception of pain and disability*	*Lifting capacity*	Yes	Inferred
Miyamoto et al. (136)	"goal of *improving disability and reducing absence from work* due to physical and functional recovery. Exercise may also *reduce pain* by influencing the endogenous inhibitory system and inducing hypoalgesia… *catastrophising and kinesiophobia* appear to be related with pain and disability… and exercise may promote benefits to improve these psychological factors."	*Pain intensity (NRS); Physical function (RMDQ)*	Yes	Specified
Moffett et al. (137)	"It may help patients overcome their *fear of physical activity* by demonstrating that movement can relieve pain; it may *reduce anxiety and depression*, and help them take *control of their situation*. These factors may enable the individual to *cope better* and return to their normal activities sooner, thus preventing *long-term disability*. "	*Activity avoidance, Physical function*	Yes	Specified
Russell et al. (138)	The aims of the programme were to encourage *normal movement*; increase participant's *confidence* in their spines; help participants *take control of their problem*.	Physical function (RMDQ)	No	Specified
Saper et al. (139)	*Improved mood, stress reduction and lower health-care costs, increases strength and flexibility.*	Physical function; Pain	No	Specified
Shirado et al. (140)	"Increasing overall *physical activity* and *spinal mobility*." Trunk muscle strengthening and flexibility.	Pain, Physical Function, Health related quality of life	No	Specified

Storro et al. (141)	"the exercises… aimed at *posture improvement*, the improvement *of aerobic capacity and strength* as well as *flexibility* of skeletal muscles related to pain perception."	Work status	No	Specified
Tilbrook et al. (142)	"intention of *improving mobility, strength*, and p*ostur*e and *reducing pain*"	Physical function (RMDQ)	No	Specified

d. Extracted Treatment Targets of Included Trials

Treatment Target	**Frequency**	**Trials**
Reduce back pain	9	Maul *et al.*, 2005; Ferreira *et al.*, 2007; Tilbrook *et al.*, 2011; Chen *et al.*, 2014; Groessl *et al.*, 2017; Harris *et al.*, 2017; Garcia *et al.*, 2018; Járomi *et al.*, 2018; Miyamoto *et al.*, 2018
Strengthening/ Muscle strength	8	Hansen *et al.*, 1993; Storrø, Moen and Svebak, 2004; Cambron *et al.*, 2006; Shirado *et al.*, 2010; Hall *et al.*, 2011; Tilbrook *et al.*, 2011; Groessl *et al.*, 2017; Saper *et al.*, 2017
Spinal stabilisation / Spinal control altered /Trunk stability	7	Maul *et al.*, 2005; Goldby *et al.*, 2006; Ferreira *et al.*, 2007; Costa *et al.*, 2009; Díaz Arribas *et al.*, 2009; Shirado *et al.*, 2010; Bronfort *et al.*, 2011
Stretching/ Flexibility	7	Storrø, Moen and Svebak, 2004; Cambron *et al.*, 2006; Shirado *et al.*, 2010; Hall *et al.*, 2011; Groessl *et al.*, 2017; Saper *et al.*, 2017; Garcia *et al.*, 2018
Posture	4	Hildebrandt *et al.*, 2000; Storrø, Moen and Svebak, 2004; Costa *et al.*, 2009; Tilbrook *et al.*, 2011
Self-confidence	3	Russell *et al.*, 2004; Maul *et al.*, 2005; Tilbrook *et al*., 2011
Functional capacity/ improve function	3	Storrø, Moen and Svebak, 2004; Maul *et al.*, 2005; Saper *et al.*, 2017
Improve physical capability / activity	3	Storrø, Moen and Svebak, 2004; Maul *et al.*, 2005; Harris *et al.*, 2017
Increase trunk muscle endurance	2	Hansen *et al.,* 1993; Bronfort *et al.*, 2011
Disability	2	Maul *et al.*, 2005; Miyamoto *et al*., 2018
Reduce stress	2	Groessl *et al*., 2017; Saper *et al.*, 2017
Recovery	2	Hildebrandt *et al.*, 2000, Cecchi *et al*., 2010
Self-efficacy; improved control	2	Russell *et al.,*, 2004; Chen *et al.*, 2014
Fear of movement associated with pain	1	Ferreira *et al.*, 2007
Reduce deconditioning	1	Ferreira *et al.*, 2007
Prevent recurrence and chronicity	1	Storrø, Moen and Svebak, 2004
Balance	1	Hall *et al.*, 2011
Body awareness	1	Hall *et al.*, 2011
General health and well-being	1	Hall *et al.*, 2011
Muscle tension	1	Hansen *et al*., 1993
Relaxation	1	Saper *et al.*, 2017
Encourage normal movement	1	Russell *et al.*, 2004
Psychosocial factors	1	Johnson *et al.*, 2007
Mobility	1	Tilbrook *et al.*, 2011
Mental positivity	1	Tilbrook *et al.*, 2011

Reduce anxiety and depression	1	Moffett *et al.*, 2006
Reduce absence from work	1	Miyamoto *et al*., 2018
Catastrophising	1	Miyamoto *et al*., 2018
Kinesiophobia	1	Miyamoto *et al*., 2018
Improved coping	1	Moffett *et al.,* 2006
Lifting capacity	1	Jaromi *et al.*, 2018

e. Extracted Secondary Outcome Domains and Measures of Included Trials

Outcome domain	Cites	Outcome Measure	Cites	Refs
HRQoL (27)				
Health Related Quality of Life	17	Short Form(SF)-36	5	Russell *et al.*, 2004; Cambron *et al.*, 2006*; Díaz Arribas *et al.*, 2009; Bronfort *et al.*, 2011; Saper *et al.*, 2017
		EuroQoL-5D(EQ-5D)	5	Russell *et al.*, 2004; Moffett *et al.*, 2006; Johnson *et al.*, 2007; Chown *et al.*, 2008; Tilbrook *et al.*, 2011
		SF-12	4	Moffett *et al.*, 2006; Albaladejo *et al.*, 2010; Tilbrook *et al.*, 2011; Groessl *et al.*, 2017
		Nottingham Health Profile	1	Goldby *et al.*, 2006
		SF6-D	1	Miyamoto *et al.*, 2018
		Self-report 4-point scale	1	Jans *et al.*, 2006
Health-care Utilisation	5	Self-report	5	Cambron *et al*., 2006; Jans *et al.,* 2006; Bronfort et al., 2011; Groessl et al., 2017; Tilbrook et al., 2011
Fatigue	2	Fatigue Severity Scale	1	Groessl *et al*., 2017
		Pittsburgh Sleep Quality Index	1	Groessl *et al.,* 2017
Dysfunction Level	1	Japan Low Back Pain Evaluation Questionnaire	1	Shirado *et al.,* 2010*
Subjective Health Complaints	1	Subjective Health Complaints Inventory	1	Harris *et al*., 2017
General Well Being	1	General Wellbeing Questionnaire	1	Maul *et al.,* 2005
Pain (26)				
Pain	19	Visual Analogue Scale	7	Ferreira *et al.*, 2007; Albaladejo *et al.*, 2010; Chen *et al.* 2014*, Jaromi *et al.,* 2018*; Cambron et al., 2006*; Johnson et al., 2007*; Shirado et al., 2010*
		Numeric Rating Scale	5	Maul *et al.*, 2005; Saper *et al.*, 2017*; Miyamoto *et al.*, 2018*; Costa et al., 2009*; Garcia et al., 2018*
		Van Korff Pain Scale	2	Russell *et al.*, 2004; Jans *et al.*, 2006
		Nordic LBP Questionnaire (frequency, duration)	1	Maul *et al.*, 2005

		Brief Pain Inventory	1	Groessl *et al.*, 2017
		Short-form McGill Questionnaire	1	Maul *et al.*, 2005
		Aberdeen Back Pain Scale	1	Tilbrook *et al.*, 2011
		Roland and Morris Pain Rating Scale	1	Cecchi *et al.*, 2010
Frequency Analgesic Use	5	Frequency	5	Maul *et al.*, 2005; Jans *et al.*, 2006; Bronfort *et al.*, 2011; Tilbrook *et al.*, 2011; Groessl et al., 2017
Pain Diagram	2	Pain Diagram	1	Goldby *et al.*, 2006
		Quantitative Pain Drawing	1	Maul *et al.*, 2005
Physical Function (26)				
Disability / Activity limitation	20	Roland Morris Disability Questionnaire	11	Maul et al., 2005; Cambron et al., 2006*; Moffett et al., 2006*; Ferreira et al., 2007; Johnson et al., 2007*; Costa et al., 2009; Shirado et al., 2010*; Bronfort et al., 2011; Hall et al., 2011; Garcia et al., 2018*; Miyamoto et al., 2018*
		Oswestry Disability Index	4	Goldby et al., 2006; Díaz Arribas et al., 2009; Harris et al., 2017; Hall et al., 2011
		Quebec Back Pain and Disability Scale	2	Jans *et al.*, 2006; Hall *et al.,* 2011
		Pain Disability Index	2	Hall *et al.*, 2011; Groessl *et al.,* 2017
		Waddell Index	1	Maul *et al.*, 2005
Activity / Function	5	Patient-Specific Functional Scale	5	Costa *et al.*, 2009*; Ferreira et al., 2007*; Hall *et al.,* 2011; Garcia *et al*., 2018; Miyamoto *et al.,* 2018;
Handicap	1	Low Back Outcome Score	1	Goldby *et al*., 2006
Beliefs/ Expectations (19)				
Fear-Avoidance Beliefs/ Kinesiophobia	5	Tampa Scale of Kinesiophobia	3	Moffett *et al.*, 2006*; Garcia *et al.*, 2017; Miyamoto *et al.*, 2018
		Fear-Avoidance Beliefs Questionnaire (work and physical activity)	2	Russell *et al*., 2004; Harris *et al.*, 2017
Self-efficacy	4	Pain Self-efficacy Questionnaire	2	Moffett *et al.*, 2006; Tilbrook *et al.*, 2011

		Exercise Self-efficacy	1	Chen *et al.,* 2014*
		6 items on 6-point Likert scale	1	Groessl *et al.*, 2017
Anxiety and Depression	4	Hospitality Anxiety and Depression Scale	2	Moffett *et al.*, 2006; Harris *et al.*, 2017
		Brief Anxiety Inventory	1	Groessl *et al.* 2017
		Centre for Epidemiologic Studies Short Depression Scale	1	Groessl *et al.* 2017
Coping/ Catastrophising	3	Utrecht Coping List from Coping and Defence Inventory	1	Harris *et al.*, 2017
		Coping Strategies Questionnaire	1	Albaladejo *et al.*, 2010
		Pain Catastrophising Scale	1	Miyamoto *et al.*, 2018
General Beliefs	2	Sense of Coherence questionnaire	1	Maul *et al.*, 2005
		Back Beliefs Questionnaire	1	Russell et al., 2004
Control	1	Multidimensional health locus of control	1	Moffett *et al.*, 2006
Physical Performance (18)				
Dynamic Motion Lumbar Spine	5	Lumbar Range (Forward Flexion Distance)	2	Goldby *et al.*, 2006; Shirado *et al.*, 2010
		Inclinometer Lumbar Range of Movement	2	Maul *et al.*, 2005 Groessl *et al.,* 2017
		CA 6000 Spine Motion Analyser	1	Bronfort *et al.*, 2011
Strengthening	5	Grip Strength (Hand Held Dynamometer)	1	Groessl *et al.* 2017
		Core Strength (Prone and Supine Bridge Positions)	1	Groessl *et al.* 2107
		Lifting capacity	1	Jaromi *et al.,* 2018
		Isometric Trunk Strength (Computerized Digital Myograph)	1	Bronfort *et al.*, 2011
		Isokinetic Strength Cybex Peak Torques	1	Maul *et al.*, 2005
Aerobic Capacity	3	Box Step Test 3 min	1	Maul *et al.*, 2005
		Shuttle walk test	1	Chown *et al.,* 2008
		Timed Walking Test	1	Goldby *et al.*, 2006
Posture	2	Qualitatively and Quantitatively	1	Hildebrandt *et al.*, 2000
		Self-reported Improvement	1	Hildebrandt *et al.*, 2000
Isometric Muscle Endurance	2	Sorensen Test and Flexion, squatting, upper arm and shoulder girdle muscles	1	Maul *et al.,* 2005
		Length of time the patient was able to maintain an unsupported upper body in the prone and supine position was recorded in seconds.	1	Bronfort *et al.*, 2011
Balance Test	1	Single Leg Stance	1	Groessl *et al.* 2017

Global Assessment (16)				
Patient Perceived Global Improvement	9	Global Perceived Effect Scale (11-point)	6	Russell et al., 2004; Ferreira *et al.*, 2007*; Costa *et al.*, 2009*; Hall *et al.*, 2011; Garcia *et al.*, 2017; Miyamoto *et al.*, 2018
		9-point Ordinal Scale	1	Bronfort *et al.*, 2011
		7-point Scale	1	Saper *et al.*, 2014
Overall Treatment Effect	2	Visual Analogue Scale	1	Hansen *et al.*, 1993
		Likert Scale change in pain/ functional capacity and satisfaction benefit of treatment	1	Maul *et al.*, 2005
Recovery	2	Pain free > 1 month	1	Costa *et al.*, 2009
		7-point scale	1	Jans *et al.*, 2006
Expectancy Improvement	2	0-10 scale	1	Garcia *et al.*, 2017
		Beliefs, expectations and preferences	1	Tilbrook, Cox and Hewitt, 2011
Recurrence	1	New episode lasting > 24 hours in those who had recovered	1	Costa *et al.*, 2009
Satisfaction (3)				
Satisfaction	3	7-point ordinal scale	1	Bronfort *et al.*, 2011
		5-point ordinal scale	1	Saper *et al.*, 2014
		4 item scale	1	Costa *et al.*, 2009

*designates where more than one primary outcome domain was designated, and the other nominated primary outcomes were then counted as secondary outcomes. Blue boxes designate most frequently cited outcome domain; Green boxes designate most frequently cited outcome measure

f. Sensitivity Analyses Using Other Values

i. Ratio of Means

Comparison of results of ratio of means with SMD values per trial and outcome

	Trial	Outcome Domain	Comparison Groups	SMD with (95% confidence intervals)	Interpretation SMD	Ratio of Means (95% confidence intervals)	Interpretation RoM
Matched	Chen *et al.,* 2014	*Pain*	*Ex vs CG*	*0.82 (0.47, 1.17)*	*Large in favour exercise*	*0.81 (0.29, 1.33)* Ψ	The mean of the intervention group is 19 % lower than the control group (medium).
	Hildebrandt *et al.,* 2000	*Recovery*	*Ex vs CG*	*0.69 (0.37, 1.0)*	*Medium-large in favour exercise*	*1.26 (0.98, 1.54)* Ψ	The mean of the intervention group is 26% higher than the control group (medium).
	Járomi *et al*., 2018	Pain	Ex vs CG	*6.50 (6.16, 6.83)*	*Very large in favour of exercise*	0.44 (0.17, 0.71) Ψ	The mean of the intervention group is 66% lower than that of the control group (very large).
	Maul *et al.,* 2005	Lifting Capacity	Ex vs CG	0.37 (-0.02, 0.76)	Small- medium in favour exercise	1.03 (0.96, 1.10) Ψ	The mean of the intervention group is 3% higher than the control group (small).
	Miyamoto *et al*., 2018	*Pain*	*Ex1 vs CG*	*0.84 (0.62,1.06)*	*Medium-large in favour exercise*	*0.86 (0.68, 1.05)* Ψ	The mean of the intervention group is 14% lower than the control group (small)
			Ex2 vs CG	*0.98 (0.80,1.16))*	*Large in favour exercise*	*0.79 (0.60, 0.99)* Ψ	*The mean of the intervention group is 21% lower than the control group (medium).*
			Ex3 vs CG	*1.30 (1.07,1.52)*	*Very large in favour of exercise*	*0.78 (0.58, 0.99)* Ψ	*The mean is 22% lower than the control group (medium).*
			Combined vs CG	*1.02 (0.77,1.27)*	*Large in favour of exercise*	*0.82 (0.70, 0.93)* Ψ	*The mean is 18% lower than the control group (small to medium).*
		Physical Function	*Ex1 vs CG*	*0.66 (0.58,0.74)*	*Small-medium in favour exercise*	*0.85 (0.66, 1.05)*	The mean is 15% lower than the control group (small to medium).
			Ex2 vs CG	*0.87 (0.79, 0.95)*	*Large in favour exercise*	*0.80 (0.59, 1.02)* Ψ	The mean is 20% lower than the control group (small to medium).
			Ex3 vs CG	*1.02 (0.94, 1.10)*	*Medium-large in favour exercise*	*0.77 (0.53, 1.00)*	The mean is 23% lower than the control group (medium).

			Combined vs CG	*0.9 (0.49,1.30)*	*Large in favour of exercise*	*0.81 (0.65, 0.97)* Ψ	*The mean is 19% lower than the control group (small to medium).*
	Moffett *et al.* 2006	*Activity Avoidance*	*Ex vs CG*	*0.14 (0.06, 0.22)*	*Small in favour exercise*	0.99 (0.92, 1.05)	The mean is 1% lower than the control group (small).
		Physical Function	*Ex vs CG*	*0.41 (0.34, 0.48)*	*Small-medium in favour exercise*	0.90 (0.73, 1.06)	The mean is 10% lower than the control group (small).
	Garcia *et al*., 2017	*Pain*	*Ex vs CG*	*0.49 (0.16, 0.81)*	*Medium in favour exercise*	0.90 (0.75, 1.06) Ψ	The mean is 10% lower than the control group (small).
		Physical Function	Ex vs CG	0.32 (-0.01,0.66)	Small-medium in favour exercise	0.91 (0.67, 1.15)	The mean is 9% lower than the control group(small).
Unmatched	Bronfort *et al.*, 2011	Pain	Ex vs SMT	0.21 (-0.07, 0.50)	Small in favour exercise	0.95 (0.81, 1.09)	The mean is 5% lower than the control group (small).
	Díaz Arribas *et al.,* 2009	*Pain*	*Ex vs PT*	*1.21 (0.86, 1.56)*	*Large in favour exercise*	*0.68 (0.37, 0.99)* Ψ	*The mean is 32% lower than the control group (medium to large).*
	Costa *et al.,* 2009	*Pain*	*Ex vs CG*	*0.49 (0.07,0.90)*	*Medium in favour exercise*	0.92 (0.79, 1.05) Ψ	The mean is 8% lower than the control group (small).
	Hall *et al.,* 2011	*Pain*	*Ex vs WL*	*0.52 (0.21, 0.83)*	*Medium in favour exercise*	0.89 (0.72, 1.06) Ψ	The mean is 11% lower than the control group (small).
	Goldby *et al.*, 2006	Pain	Ex vs CG	0.24 (-0.07, 0.54)	Small-medium in favour exercise	0.91 (0.64, 1.19)	The mean is 9% lower than the control group (small).
	Hansen *et al*., 1993	Pain	Ex vs CG	0.18 (-0.95, 0.02)	Small favours exercise	0.95 (0.67, 1.22)	The mean is 5% lower than the control group (small).
	Johnson *et al.,* 2007	*Pain*	*Ex vs CG*	*0.30 (0.04, 0.57)*	*Small–medium in favour exercise*	0.92 (0.75, 1.09)	The mean is 8% lower than the control group (small).
	Saper *et al.,* 2017	Pain	Ex vs PT	0.21 (-0.09, 0.50)	Small favours exercise	0.95 (0.82, 1.08)	The mean is 5% lower than the control group (small).
	Shirado *et al*., 2010	Pain	Ex vs CG	0.18 (-0.12,0.47)	Small in favour exercise	0.90 (0.60, 1.20)	The mean is 10% lower than the control group (small).
		HRQoL	*Ex vs CG*	*0.29 (0.00, 0.57)*	*Small-medium in favour exercise*	0.80 (0.24, 1.36)	The mean is 20% lower than the control group (small to medium).
	Albaladejo *et al.,* 2010	*Physical Function*	*Ex vs Combined CG*	*0.24 (0.01,0.47)*	*Small in favour of exercise*	0.92 (0.74, 1.09)	The mean is 8% lower than the control group (small).
	Cambron *et al.*, 2006	Physical Function	*Ex vs SMT*	*-0.38 (-0.64, -0.12)*	***Medium in favour SMT***	***1.25 (0.89,1.76)***	*The mean is 25% lower in the control group (medium).*
	Cecchi *et al.*, 2010	Physical Function	*GE vs SMT*	*-0.96 (-1.09, -0.82)*	***Large in favour SMT***	***1.46 (1.05, 1.88)****	*The mean is 46% higher than the control group (very large).*

	Costa *et al.,* 2009	*Physical Function*	*Ex vs CG*	*0.63 (0.20, 1.05)*	*Medium–large in favour exercise*	1.11 (0.95, 1.27) Ψ	The mean is 11% higher than the control group (small).
	Ferreira *et al.,* 2007	*Physical Function*	*GE vs SMT*	*-0.70 (-1.19, -0.22)*	***Medium- large in favour SMT***	**0.92 (0.78, 1.05)** Ψ	The mean is 8% lower than the control group (small).
			MCE vs SMT	0.05 (-0.44, 0.53)	Small in favour MCE	1.00 (0.89, 1.12)	No difference
			Combined Int vs SMT	***-0.33 (-0.61, -0.06)***	***Small in favour SMT***	**0.96 (0.85, 1.07)** Ψ	The mean is 4% lower than the intervention group (small).
	Johnson *et al.*, 2007	Physical Function	Ex vs CG	0.15 (-0.41, 0.11)	Small in favour exercise	0.97 (0.79, 1.15)	The mean is 3% lower than the control group (small).
	Saper *et al.,* 2017	Physical Function	Ex vs PT	0.25 (-0.05, 0.54)	Small in favour exercise	0.95 (0.83, 1.08)	The mean is 5% lower than the control group (small).
	Russell *et al.*, 2004	Physical Function	*Ex vs CG*	*0.33 (0.28,0.39)*	*Small-medium in favour exercise*	0.91 (0.90,0.92)	*The mean is 9% lower than the control group (small).*
			Ex vs SMT	***-0.09 (-0.04, -0.15)***	***Small in favour of SMT***	***1.03 (1.02, 1.04)***	*The mean of the intervention is 3% higher than the control group (small).*
			Ex vs Combined CG	*0.11 (0.07, 0.15)*	*Very small in favour of exercise*	0.97 (0.85, 1.10)	The mean is 3% lower than the control group (small).
	Jans *et al.,* 2006	Recovery	Ex vs UC	0.30 (-0.06,0.66)	Small-medium in favour exercise	1.06 (0.90, 1.22)	The mean is 6% higher than the control group (small).
	Harris *et al*., 2017	Sick leave	Ex vs UC	-0.16 (-0.48,0.16)	**Small in favour control**	**0.94 (0.65, 1.23)**	The mean is 6% lower than the control group (small).
	Storrø *et al.,* 2004	*Sick leave*	*Ex vs CG*	*0.74 (0.55, 0.93)*	*Medium-large in favour of exercise*	*1.27 (1.11, 1.42)*	*The mean is 27% higher than the control group (medium).*

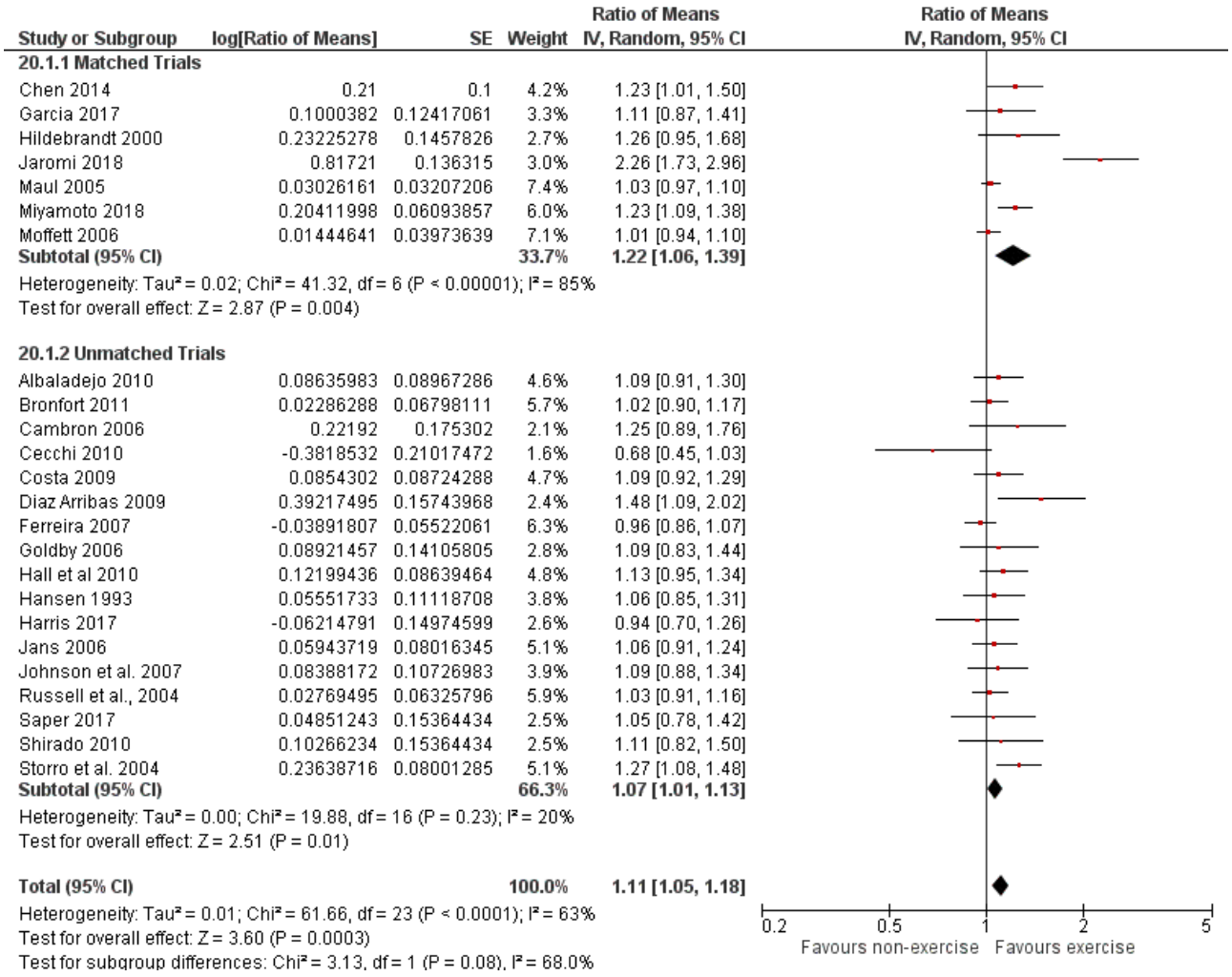
Ψ represents RoM values that differ by being one or more groups smaller than the SMD interpretation according to Cohen (1992)'s classification of small <0.2; medium 0.5 and large effect sizes >0.8 as recommended by Friedrich et al., (2011) and Fu et al., (2013). * represents RoM values that are greater than the SMD interpretation using Cohen (1992)'s interpretation as seen above. Values in italics are statistically significant. Interpretations highlighted in bold favour the control group. 95% confidence intervals are bracketed alongside SMD. The outlier is shaded in grey. Chown *et al*., 2008; Tilbrook *et al*., 2011; Groessl *et al.,* 2017 only reported change scores which prevented transformation into RoM values.

ii. Sensitivity Analysis Using Follow-up Standard Deviations

Trial	Outcome	SMD (using BL values)	SMD (FU values)
Albaladejo *et al.* 2010	PF	0.24 (0.01, 0.47)	0.12 (-0.11,0.35)
Bronfort *et al.* 2011	Pain	0.21 (-0.07,0.50)	0.15 (-0.13,0.43)
Cecchi *et al.* 2010	PF	**-0.81 (-0.91, -0.70)**	**-0.76 (-0.91, -0.70)**
Chen *et al.* 2014	Pain	**0.68 (0.33,1.02)**	**0.82 (0.47,1.17)**
Costa *et al.* 2009	Pain	0.49 (0.17,0.80)	0.37 (0.05,0.69)
Diaz Arribas *et al.* 2009	Pain	**1.21 (0.86, 1.56)**	**1.40 (1.05,1.75)**
Ferreira *et al.* 2007	PF	-0.33 (-0.61, -0.06)	-0.22 (-0.50,0.05)
Garcia *et al.* 2017	Pain	0.49 (0.16, 0.81)	0.32 (0.00,0.64)
Goldby *et al*. 2006	Pain	0.24 (-0.07,0.54)	0.17 (-0.24, 0.58)
Groessl *et a*l. 2017	PF	**0.14 (-0.18,0.46)**	**0.16 (-0.16,0.48)**
Hall *et al.* 2011	Pain	0.52 (0.21,0.83)	0.52 (0.21, 0.83)
Hansen *et al*. 1993	Pain	**0.19 (-0.40, 0.20)**	**0.24 (-0.40, 0.20)**
Johnson *et al*. 2007	Pain	**0.30 (0.04,0.57)**	**0.34 (0.07,0.62)**
Maul *et al.,* 2005	LC	0.37 (-0.02,0.76)	0.24 (-0.15,0.63)
Miyamoto *et al.* 2018	Pain	1.02 (0.77,1.27)	0.98 (0.73,1.22)
Russell *et al.* 2004	PF	0.08 (-0.07,0.23)	0.08 (-0.07, 0.23)
Saper *et al*. 2017	PF	**0.25 (-0.03, 0.52)**	**0.26 (-0.04,0.56)**
Shirado *et al.* 2010	Pain	**0.18 (-0.12, 0.47)**	**0.24 (-0.06, 0.53)**
Tilbrook *et al.* 2011	PF	0.5 (0.26,0.74)	0.41 (0.17, 0.65)

Bold values demonstrate SMD calculations that increased when using the FU values. () indicate 95% confidence intervals; SMD is standardised mean difference, BL is baseline, FU is follow-up. Harris, Hildebrandt, Jans and Storro were not included as they reported proportions, for which the standard deviation was calculated using the follow up data. Moffett et al., 2006 did not provide enough information for this to be calculated. Hansen et al. 1993, Shirado et al., 2010 and Albaladejo et al., 2010 values were calculated using median, lower quartile range and upper quartile range values.

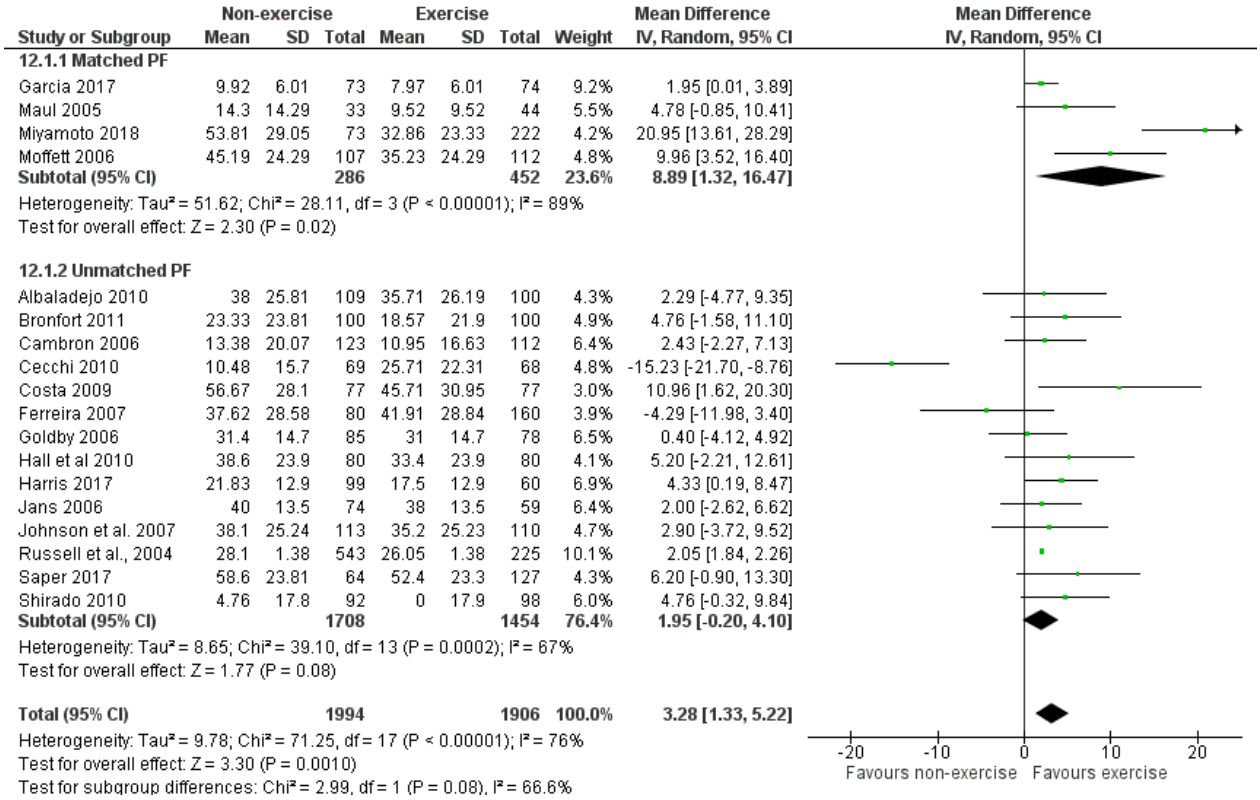
iii. Meta-Analysis Using Ratio of Means



Where SE is standard error, IV is inverse variance, CI is confidence interval. Two trials were excluded from this analysis (Tilbrook et al., 2011, Groessl et al., 2017) as they only reported change scores in their published trial reports.

iv. Effectiveness Sensitivity Analysis

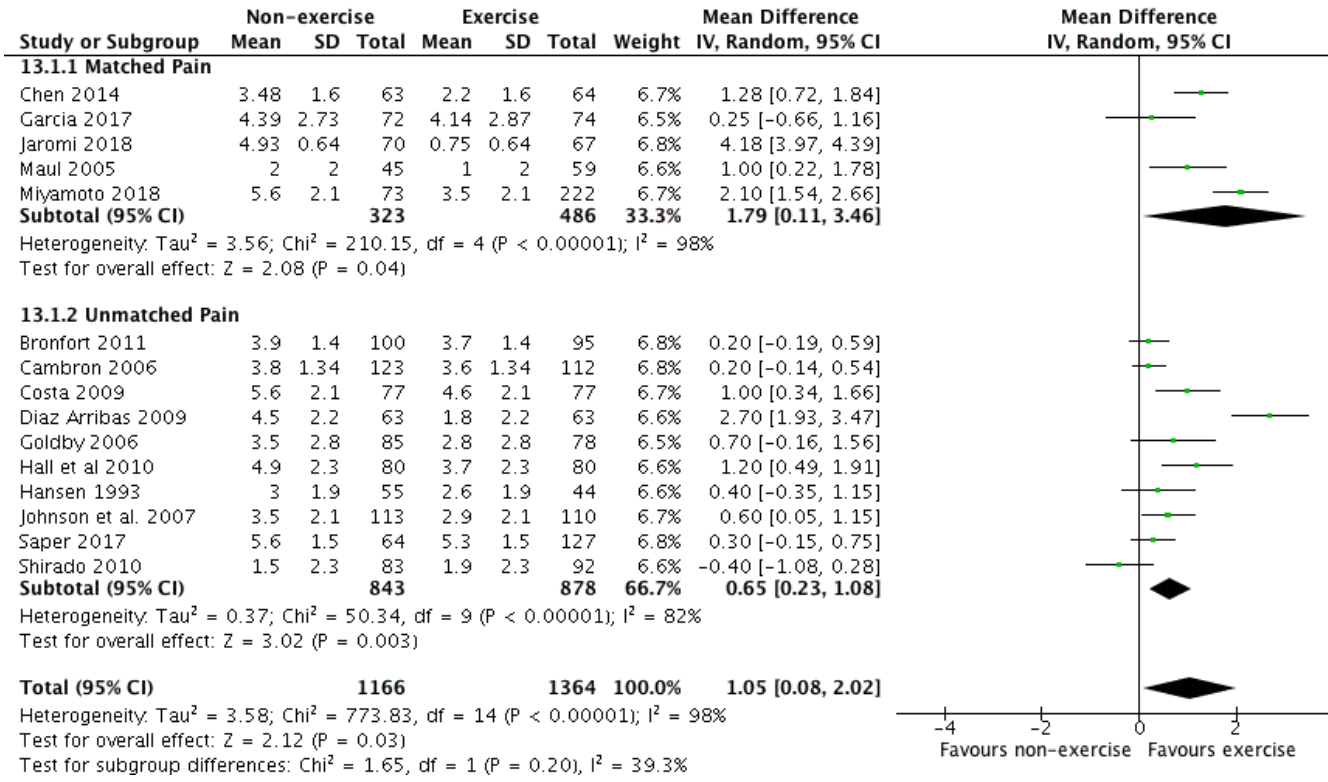
1. Physical Function



Study or Subgroup	Non-exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
12.1.1 Matched PF								
Garcia 2017	9.92	6.01	73	7.97	6.01	74	9.2%	1.95 [0.01, 3.89]
Maul 2005	14.3	14.29	33	9.52	9.52	44	5.5%	4.78 [-0.85, 10.41]
Miyamoto 2018	53.81	29.05	73	32.86	23.33	222	4.2%	20.95 [13.61, 28.29]
Moffett 2006	45.19	24.29	107	35.23	24.29	112	4.8%	9.96 [3.52, 16.40]
Subtotal (95% CI)			**286**			**452**	**23.6%**	**8.89 [1.32, 16.47]**
Heterogeneity: Tau² = 51.62; Chi² = 28.11, df = 3 (P < 0.00001); I² = 89%								
Test for overall effect: Z = 2.30 (P = 0.02)								
12.1.2 Unmatched PF								
Albaladejo 2010	38	25.81	109	35.71	26.19	100	4.3%	2.29 [-4.77, 9.35]
Bronfort 2011	23.33	23.81	100	18.57	21.9	100	4.9%	4.76 [-1.58, 11.10]
Cambron 2006	13.38	20.07	123	10.95	16.63	112	6.4%	2.43 [-2.27, 7.13]
Cecchi 2010	10.48	15.7	69	25.71	22.31	68	4.8%	-15.23 [-21.70, -8.76]
Costa 2009	56.67	28.1	77	45.71	30.95	77	3.0%	10.96 [1.62, 20.30]
Ferreira 2007	37.62	28.58	80	41.91	28.84	160	3.9%	-4.29 [-11.98, 3.40]
Goldby 2006	31.4	14.7	85	31	14.7	78	6.5%	0.40 [-4.12, 4.92]
Hall et al 2010	38.6	23.9	80	33.4	23.9	80	4.1%	5.20 [-2.21, 12.61]
Harris 2017	21.83	12.9	99	17.5	12.9	60	6.9%	4.33 [0.19, 8.47]
Jans 2006	40	13.5	74	38	13.5	59	6.4%	2.00 [-2.62, 6.62]
Johnson et al. 2007	38.1	25.24	113	35.2	25.23	110	4.7%	2.90 [-3.72, 9.52]
Russell et al., 2004	28.1	1.38	543	26.05	1.38	225	10.1%	2.05 [1.84, 2.26]
Saper 2017	58.6	23.81	64	52.4	23.3	127	4.3%	6.20 [-0.90, 13.30]
Shirado 2010	4.76	17.8	92	0	17.9	98	6.0%	4.76 [-0.32, 9.84]
Subtotal (95% CI)			**1708**			**1454**	**76.4%**	**1.95 [-0.20, 4.10]**
Heterogeneity: Tau² = 8.65; Chi² = 39.10, df = 13 (P = 0.0002); I² = 67%								
Test for overall effect: Z = 1.77 (P = 0.08)								
Total (95% CI)			**1994**			**1906**	**100.0%**	**3.28 [1.33, 5.22]**
Heterogeneity: Tau² = 9.78; Chi² = 71.25, df = 17 (P < 0.00001); I² = 76%								
Test for overall effect: Z = 3.30 (P = 0.0010)								
Test for subgroup differences: Chi² = 2.99, df = 1 (P = 0.08), I² = 66.6%								

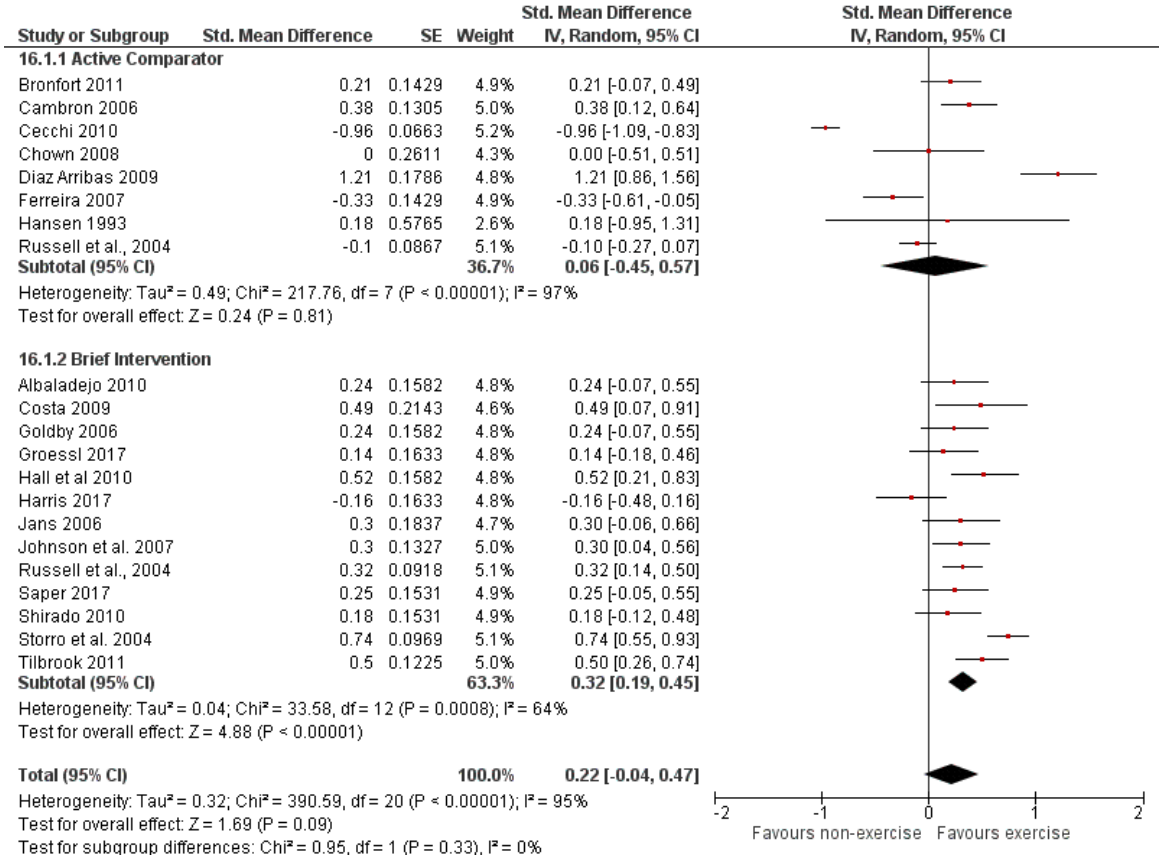
Where SE is standard error, IV is inverse variance, CI is confidence interval.

2. Pain


Non-exercise
Exercise
Mean Difference
Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
13.1.1 Matched Pain
Chen 2014 3.48 1.6 63 2.2 1.6 64 6.7% 1.28 [0.72, 1.84]
Garcia 2017 4.39 2.73 72 4.14 2.87 74 6.5% 0.25 [-0.66, 1.16]
Jaromi 2018 4.93 0.64 70 0.75 0.64 67 6.8% 4.18 [3.97, 4.39]
Maul 2005 2 2 45 1 2 59 6.6% 1.00 [0.22, 1.78]
Miyamoto 2018 5.6 2.1 73 3.5 2.1 222 6.7% 2.10 [1.54, 2.66]
Subtotal (95% CI) 323 486 33.3% 1.79 [0.11, 3.46]
Heterogeneity: Tau² = 3.56; Chi² = 210.15, df = 4 (P < 0.00001); I² = 98%
Test for overall effect: Z = 2.08 (P = 0.04)
13.1.2 Unmatched Pain
Bronfort 2011 3.9 1.4 100 3.7 1.4 95 6.8% 0.20 [-0.19, 0.59]
Cambron 2006 3.8 1.34 123 3.6 1.34 112 6.8% 0.20 [-0.14, 0.54]
Costa 2009 5.6 2.1 77 4.6 2.1 77 6.7% 1.00 [0.34, 1.66]
Diaz Arribas 2009 4.5 2.2 63 1.8 2.2 63 6.6% 2.70 [1.93, 3.47]
Goldby 2006 3.5 2.8 85 2.8 2.8 78 6.5% 0.70 [-0.16, 1.56]
Hall et al 2010 4.9 2.3 80 3.7 2.3 80 6.6% 1.20 [0.49, 1.91]
Hansen 1993 3 1.9 55 2.6 1.9 44 6.6% 0.40 [-0.35, 1.15]
Johnson et al. 2007 3.5 2.1 113 2.9 2.1 110 6.7% 0.60 [0.05, 1.15]
Saper 2017 5.6 1.5 64 5.3 1.5 127 6.8% 0.30 [-0.15, 0.75]
Shirado 2010 1.5 2.3 83 1.9 2.3 92 6.6% -0.40 [-1.08, 0.28]
Subtotal (95% CI) 843 878 66.7% 0.65 [0.23, 1.08]
Heterogeneity: Tau² = 0.37; Chi² = 50.34, df = 9 (P < 0.00001); I² = 82%
Test for overall effect: Z = 3.02 (P = 0.003)
Total (95% CI) 1166 1364 100.0% 1.05 [0.08, 2.02]
Heterogeneity: Tau² = 3.58; Chi² = 773.83, df = 14 (P < 0.00001); I² = 98%
Test for overall effect: Z = 2.12 (P = 0.03)
Test for subgroup differences: Chi² = 1.65, df = 1 (P = 0.20), I² = 39.3%
-4
-2
0
2
Favours non-exercise
Favours exercise

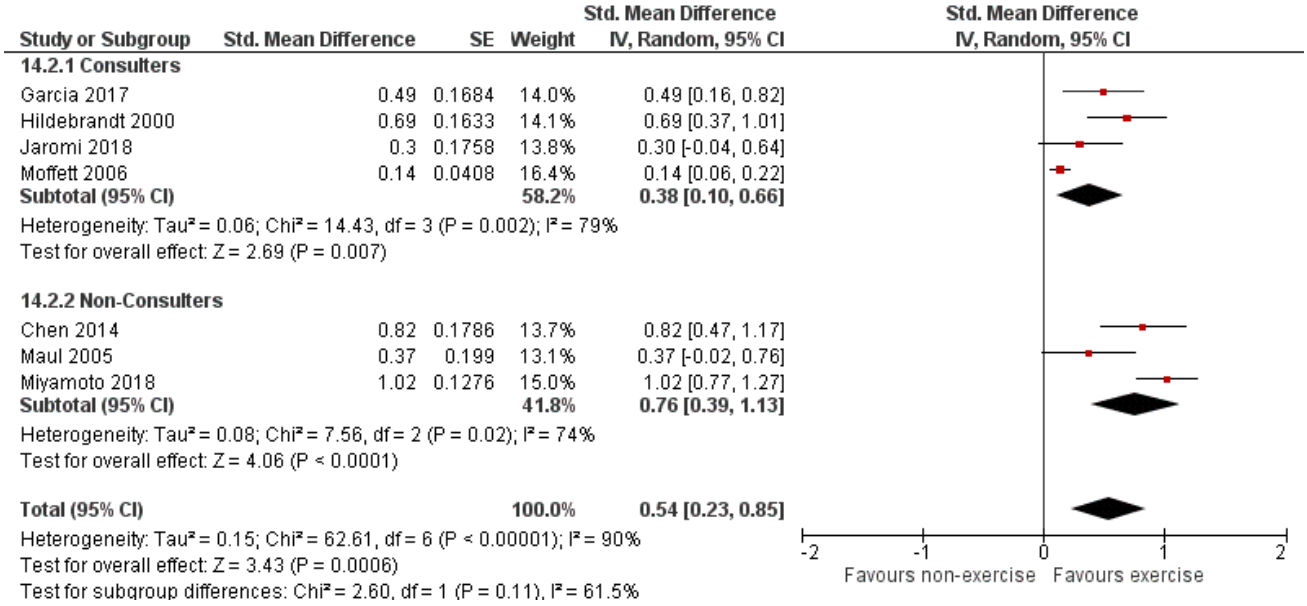
v. Sub-group Analyses

1. Sub-group Analysis of Comparator Groups in the Unmatched Group

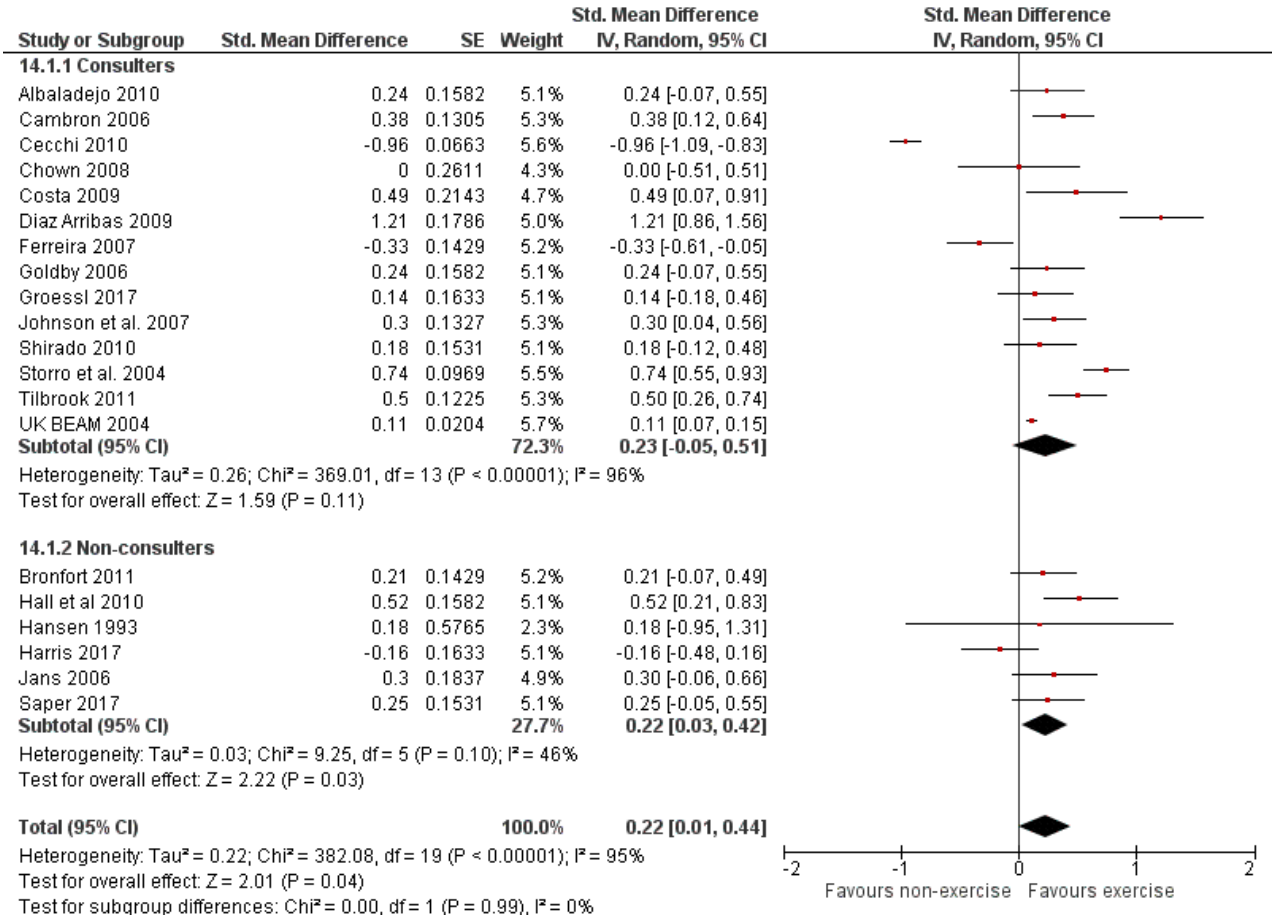


2. Sub-group Analysis of Recruitment Strategy

Matched trials:

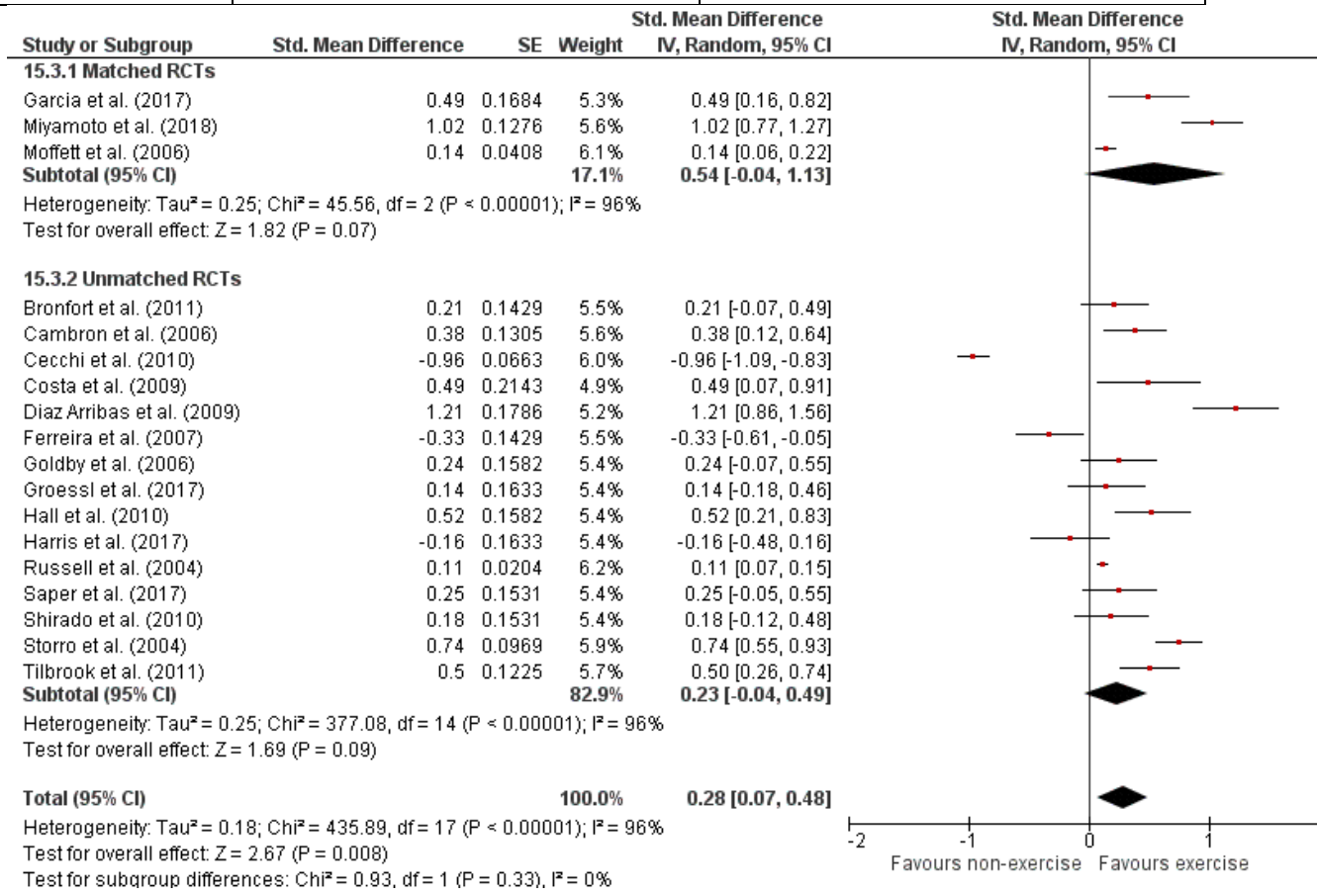


Unmatched Trials



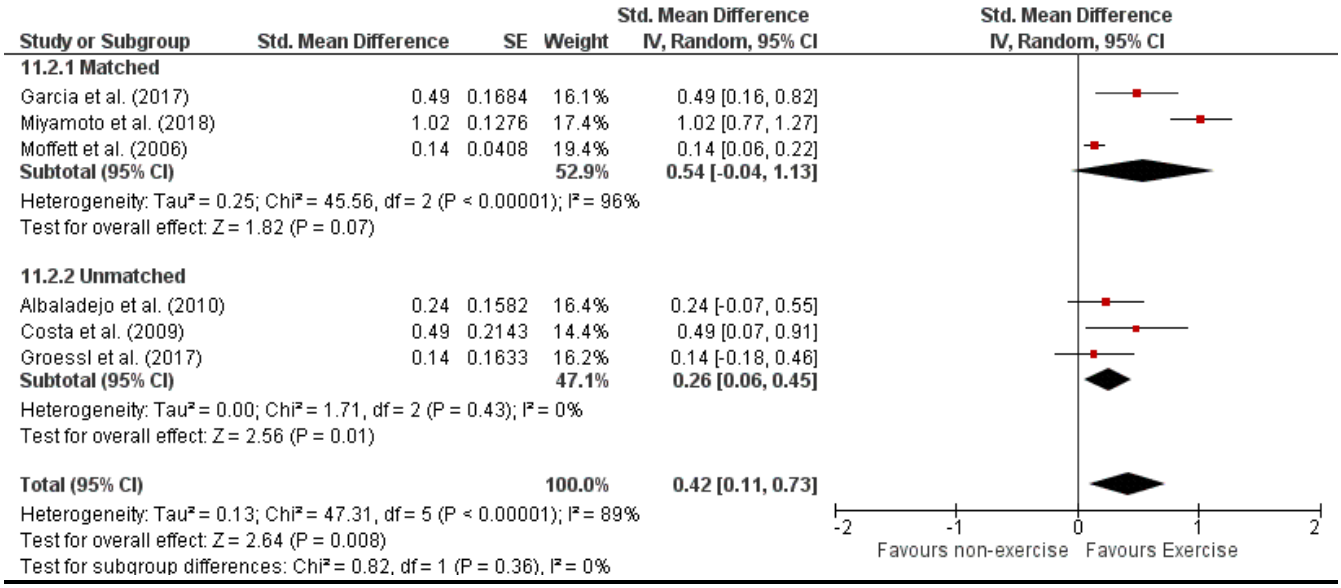
3. Sub-group Analysis of Specified Treatment Targets

Category	SMD of RCTs including both specified and inferred treatment targets	SMDs of RCTs that specified treatment targets ONLY
Matched	0.54 (95% CI 0.23 to 0.85); p=0.0006† n=1197; 7 trials	0.54 (95% CI -0.04, 1.13); p=0.07; n=662; 3 trials
Unmatched	0.22 (95% CI 0.01,0.44); p=0.04†; n=4510; 20 trials	0.23 (95% CI -0.04, 0.49) p=0.09; n=3549; 14 trials
Overall effect	0.31 (95% CI 0.01, 0.44); p=0.0002; n=5707; 27 trials;	0.28 (0.07, 0.48); p=0.008; n=4211; 17 trials
Test for Sub-group Differences	P=0.10	P=0.33



4. Sub-group Analysis of Risk of Bias

Category	Category SMD	Low Risk of Bias
Matched	0.54 (95% CI 0.23 to 0.85); p=0.0006† n=1197; 7 trials	0.54 (95% CI -0.04, 1.13); p=0.07 n=662; 3 trials
Unmatched	0.22 (95% CI 0.01,0.44); p=0.04†; n=4510; 20 trials	0.26 (95% CI 0.06, 0.45); p= 0.01; n=652; 3 trials
Overall effect	0.31 (95% CI 0.01, 0.44); p=0.0002; n=5707; 27 trials;	0.42 (95% CI 0.11, 0.73); p=0.36; n=1314; 3 trials
Test for Sub-group Differences	P=0.10	P=0.36



5. Sample size analysis across arms of included trials

Trial Lead Author	**Sample Intervention Arm**	**Sample Control Arm**
Albaladejo	*100*	*248*
Bronfort	95	100
Cambron	*112*	*123*
Cecchi	*136*	*69*
Chen	64	63
Chown	*24*	*39*
Costa	77	77
Diaz Arribas	63	63
Ferreira	74	77
Garcia	74	73
Goldby	78	85
Groessl	75	75
Hall	80	80
Hansen	*44*	*55*
Harris	*60*	*99*
Hildebrandt	72	65
Jans	*59*	*74*
Jaromi	67	70
Johnson	110	113
Maul	*59*	*45*
Miyamoto	*222*	*73*
Moffett	112	107
Russell	*225*	*543*
Saper	*127*	*64*
Shirado	*92*	*83*
Storro	200	200
Tilbrook	135	139

Italic figures denote where sample sizes differed more than 10% between arms.

g. Data Sharing Agreement for Datasets used in Chapter 4

Research Institute for Primary Care and Health Sciences (iPCHS) and Keele CTU, Keele University

Record of external data stored at the iPCHS / Keele CTU provided by an external source (e.g. another university or NHS organisation)

To be completed by the Keele Researcher of the proposed study

The lead iPCHS / CTU researcher should complete this form in conjunction with the Principal Investigator (PI) of the collaborating unit which is providing the external data to be stored at Keele iPCHS / CTU

Proposed Study Title: *for which data is being collected*	Exercise, as a complex intervention for chronic low back pain – does matching the primary outcomes to treatment targets matter?	
Collaborative Research? *Pease tick ✓ selection*		**No** collaboration, data/data provision to Keele with no collaboration

Keele Study Team:

This section of the form will include members of the iPCHS / CTU and will be used to support decisions on authorship of papers generated from this secondary analysis (if applicable)

Details	*Name*	*Role in study team*	*Contact details*
	L Wood	PhD Student	L.wood2@keele.ac.uk
	A Bishop	Lead Supervisor	a.bishop@keele.ac.uk
	N Foster	Supervisor	n.foster@keele.ac.uk
	M Lewis	Supervisor	a.m.lewis@keele.ac.uk

Institution/Organisation providing data:

Please provide details of the data owner/custodian and personnel providing approval of data release.

External Research Team Details	*Name*	*Role in study team*	*Contact details*
Integrative Health & Wellbeing Research Program Earl E. Bakken Center for Spirituality & Healing University of Minnesota	Gert Bronfort	Lead Author	bronf003@umn.edu

Details of study providing dataset:

Study name	**Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: a randomized clinical trial**
Sponsor	Wolfe Harris Centre for Clinical Studies, Northwestern Health Sciences University USA
Funder	None
Funder Reference No.	*None*

Proposed Study Details

Research Question / Objective: 1) Does a composite primary outcome matched to the treatment targets of the intervention change the conclusion(s) of the RCT? 2) A secondary analysis aims to test the hypothesis that generating a composite primary outcome that includes the matched treatment targets of exercise in NSCLBP RCTs alters the conclusion of RCTs.	
Analysis planned Start Date: October 2018	**End date: June 2019**

In to which iPCHS programme(s) does this study fit? *(Please tick all that apply)*

Osteoporosis and Osteoarthritis		Prognosis and Consultation Epidemiology	
Musculoskeletal Health and Pain	x	Education & Training	
Inflammatory studies		Secondary Care	
Mental Health studies		Other (please describe)	

Brief summary of process gone through to gain approval to hold data: *Attach copies of supporting documents submitted to external institution and confirmation of approval (e.g. data sharing agreement.)* Please see attached protocol				
Are there any specific conditions applied to the use of the data by the collaborating site:	**YES**		**NO**	**x**
If yes, please specify:				
Details of any requirements for destroying data at end of study: **None provided**				
Type of data provided: Anonymous dataset of 7 variables at all follow up time points				
Does the data set contain any personally identifiable data?	**YES**		**NO**	**x**
If yes, please specify:				
Expected server path / file location to hold data at Keele: **Pre87 S:\CHp 5 exploratory analysis**				

Keele iPCHS / CTU Lead Researcher: please complete the form and email a copy of completed form and provide signed a hard copy to the Keele CTU Data Systems Manager).

Please note on receiving this data into the iPCHS / Keele CTU you (Keele PI) are agreeing to:

- adhere to the Research Institute for Primary Care and Health Sciences (iPCHS) and CTU (where appropriate) data management policy, data security and applicable procedures.
- take full responsibility for the data held in the centre.
- comply with the principles and conditions set out in the Data Protection Act 1998, the Research Governance Framework and maintain proper safeguards to ensure confidentiality.
- only use the data for the approved purpose of the release and not allow further data release
- satisfy acknowledgement, co authorship and publication permissions (if applicable)

Name of Keele PI	Signature	Date
L Wood	*L Wood*	5/11/2018

Research Institute for Primary Care and Health Sciences (iPCHS) and Keele CTU, Keele University

Record of external data stored at the iPCHS / Keele CTU provided by an external source (e.g. another university or NHS organisation)

To be completed by the Keele Researcher of the proposed study

The lead iPCHS / CTU researcher should complete this form in conjunction with the Principal Investigator (PI) of the collaborating unit which is providing the external data to be stored at Keele iPCHS / CTU

Proposed Study Title: *for which data is being collected*	Exercise, as a complex intervention for chronic low back pain – does matching the primary outcomes to treatment targets matter?	
Collaborative Research? *Pease tick ✓ selection*		**No** collaboration, data/data provision to Keele with no collaboration

Keele Study Team:

This section of the form will include members of the iPCHS / CTU and will be used to support decisions on authorship of papers generated from this secondary analysis (if applicable)

Details	***Name***	***Role in study team***	***Contact details***
	L Wood	PhD Student	L.wood2@keele.ac.uk
	A Bishop	Lead Supervisor	a.bishop@keele.ac.uk
	N Foster	Supervisor	n.foster@keele.ac.uk
	M Lewis	Supervisor	a.m.lewis@keele.ac.uk

Institution/Organisation providing data:

Please provide details of the data owner/custodian and personnel providing approval of data release.

External Research Team Details	***Name***	***Role in study team***	***Contact details***
Department of Family Medicine and Public Health, University of California, San Diego, California	Erik Groessl	Lead Author	egroessl@ucsd.edu

Details of study providing dataset:

Study name	Yoga for Military Veterans with Chronic Low Back Pain: A Randomized Clinical Trial
Sponsor	*Department of Family Medicine and Public Health, University of California, San Diego, California*
Funder	This research was funded by a grant from Veteran Affairs Rehabilitation Research and Development
Funder Reference No.	**Grant #RX000474**

Proposed Study Details

Research Question / Objective:

1. The primary analysis aims to test the hypothesis that matching the primary outcome measures to the reported treatment targets of exercise in NSCLBP RCTs alters the key conclusion of RCTs.
2. Does a composite primary outcome matched to the treatment targets of the intervention change the conclusion(s) of the RCT?

Analysis planned Start Date: October 2018	**End date: June 2019**

In to which iPCHS programme(s) does this study fit? *(Please tick all that apply)*

Osteoporosis and Osteoarthritis		**Prognosis and Consultation Epidemiology**	
Musculoskeletal Health and Pain	x	**Education & Training**	
Inflammatory studies		**Secondary Care**	
Mental Health studies		**Other (please describe)**	

Brief summary of process gone through to gain approval to hold data:
Attach copies of supporting documents submitted to external institution and confirmation of approval (e.g. data sharing agreement.)
Please see attached protocol

Are there any specific conditions applied to the use of the data by the collaborating site:	**YES**		**NO**	**x**
If yes, please specify:				

Details of any requirements for destroying data at end of study:
None provided

Type of data provided:
Anonymous dataset of 5 variables at one follow up time point

Does the data set contain any personally identifiable data?	**YES**		**NO**	**x**
If yes, please specify:				

Expected server path / file location to hold data at Keele:

Pre87 S:\CHp 5 exploratory analysis

Keele iPCHS / CTU Lead Researcher: please complete the form and email a copy of completed form and provide signed a hard copy to the Keele CTU Data Systems Manager).

Please note on receiving this data into the iPCHS / Keele CTU you (Keele PI) are agreeing to:

- adhere to the Research Institute for Primary Care and Health Sciences (iPCHS) and CTU (where appropriate) data management policy, data security and applicable procedures.
- take full responsibility for the data held in the centre.
- comply with the principles and conditions set out in the Data Protection Act 1998, the Research Governance Framework and maintain proper safeguards to ensure confidentiality.
- only use the data for the approved purpose of the release and not allow further data release
- satisfy acknowledgement, co authorship and publication permissions (if applicable)

Name of Keele PI	**Signature**	**Date**
L Wood	*L Wood*	**12/10/2018**

Research Institute for Primary Care and Health Sciences (iPCHS) and Keele CTU, Keele University

Record of external data stored at the iPCHS / Keele CTU provided by an external source (e.g. another university or NHS organisation)

To be completed by the Keele Researcher of the proposed study

The lead iPCHS / CTU researcher should complete this form in conjunction with the Principal Investigator (PI) of the collaborating unit which is providing the external data to be stored at Keele iPCHS / CTU

Proposed Study Title: *for which data is being collected*	Exercise, as a complex intervention for chronic low back pain – does matching the primary outcomes to treatment targets matter?	
Collaborative Research? *Pease tick ✓ selection*		**No** collaboration, data/data provision to Keele with no collaboration

Keele Study Team:

This section of the form will include members of the iPCHS / CTU and will be used to support decisions on authorship of papers generated from this secondary analysis (if applicable)

Details	*Name*	*Role in study team*	*Contact details*
	L Wood	PhD Student	L.wood2@keele.ac.uk
	A Bishop	Lead Supervisor	a.bishop@keele.ac.uk
	N Foster	Supervisor	n.foster@keele.ac.uk
	M Lewis	Supervisor	a.m.lewis@keele.ac.uk

Institution/Organisation providing data:

Please provide details of the data owner/custodian and personnel providing approval of data release.

External Research Team Details	*Name*	*Role in study team*	*Contact details*
Data custodian University of Oslo, Norway	Silje Endersen Reme	Author	Silje.reme@psukologi.uio.no

Details of study providing dataset:

Study name	Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (the CINS trial)
Sponsor	University of Oslo
Funder	This research project was funded by The Research Council of Norway and has also been financially supported by the Norwegian ExtraFoundation for Health and Rehabilitation through EXTRA funds
Funder Reference No.	(175466/V50)

Proposed Study Details

Research Question / Objective:
Does an outcome matched to the treatment targets of the intervention produce greater effect sizes and increased statistical significance than the primary outcome, and does it change the conclusion(s) of the RCT?

Analysis planned Start Date: October 2018	**End date: Feburary 2019**

In to which iPCHS programme(s) does this study fit? *(Please tick all that apply)*

Osteoporosis and Osteoarthritis		**Prognosis and Consultation Epidemiology**	
Musculoskeletal Health and Pain	x	**Education & Training**	
Inflammatory studies		**Secondary Care**	
Mental Health studies		**Other (please describe)**	

Brief summary of process gone through to gain approval to hold data:
Attach copies of supporting documents submitted to external institution and confirmation of approval (e.g. data sharing agreement.)
Please see attached protocol

Are there any specific conditions applied to the use of the data by the collaborating site:	**YES**		**NO**	**x**
If yes, please specify:				

Details of any requirements for destroying data at end of study:
None provided

Type of data provided:
Anonymous dataset of 9 variables at one follow up time point and baseline.

Does the data set contain any personally identifiable data?	**YES**		**NO**	**x**
If yes, please specify:				

Expected server path / file location to hold data at Keele:

Pre87 S:\CHp 5 exploratory analysis

Keele iPCHS / CTU Lead Researcher: please complete the form and email a copy of completed form and provide signed a hard copy to the Keele CTU Data Systems Manager).

Please note on receiving this data into the iPCHS / Keele CTU you (Keele PI) are agreeing to:

- adhere to the Research Institute for Primary Care and Health Sciences (iPCHS) and CTU (where appropriate) data management policy, data security and applicable procedures.
- take full responsibility for the data held in the centre.
- comply with the principles and conditions set out in the Data Protection Act 1998, the Research Governance Framework and maintain proper safeguards to ensure confidentiality.
- only use the data for the approved purpose of the release and not allow further data release
- satisfy acknowledgement, co authorship and publication permissions (if applicable)

Name of Keele PI	**Signature**	**Date**
L Wood	*L Wood*	**16-1-2019**

h. Data Sharing Agreement for Datasets used in Chapter 5

Research Institute for Primary Care and Health Sciences (iPCHS) and Keele CTU, Keele University

Record of external data stored at the iPCHS / Keele CTU provided by an external source (e.g. another university or NHS organisation)

To be completed by the Keele Researcher of the proposed study

The lead iPCHS / CTU researcher should complete this form in conjunction with the Principal Investigator (PI) of the collaborating unit which is providing the external data to be stored at Keele iPCHS / CTU

Proposed Study Title: *for which data is being collected*	Exercise, as a complex intervention for chronic low back pain – does matching the primary outcomes to treatment targets matter?	
Collaborative Research? *Pease tick ✓ selection*		**No** collaboration, data/data provision to Keele with no collaboration

Keele Study Team:

This section of the form will include members of the iPCHS / CTU and will be used to support decisions on authorship of papers generated from this secondary analysis (if applicable)

Details	*Name*	*Role in study team*	*Contact details*
	L Wood	PhD Student	L.wood2@keele.ac.uk
	A Bishop	Lead Supervisor	a.bishop@keele.ac.uk
	N Foster	Supervisor	n.foster@keele.ac.uk
	M Lewis	Supervisor	a.m.lewis@keele.ac.uk

Institution/Organisation providing data:

Please provide details of the data owner/custodian and personnel providing approval of data release.

External Research Team Details	*Name*	*Role in study team*	*Contact details*
Data custodian York, UK	David Torgerson	Author, Director of Clinical Trials Unit	david.torgerson@york.ac.uk

Details of study providing dataset:

Study name	Randomized trial of two physiotherapy interventions for primary care neck and back pain patients: McKenzie vs brief physiotherapy pain management
Sponsor	University of Hull
Funder	Arthritis Research Campaign
Funder Reference No.	

Proposed Study Details

Research Question / Objective:
Does a composite primary outcome matched to the treatment targets of the intervention change the conclusion(s) of the RCT?

Analysis planned Start Date: October 2018	**End date: June 2019**

In to which iPCHS programme(s) does this study fit? *(Please tick all that apply)*

Osteoporosis and Osteoarthritis		**Prognosis and Consultation Epidemiology**	
Musculoskeletal Health and Pain	x	**Education & Training**	
Inflammatory studies		**Secondary Care**	
Mental Health studies		**Other (please describe)**	

Brief summary of process gone through to gain approval to hold data:
Attach copies of supporting documents submitted to external institution and confirmation of approval (e.g. data sharing agreement.)
Please see attached protocol

Are there any specific conditions applied to the use of the data by the collaborating site:	**YES**		**NO**	**x**
If yes, please specify:				

Details of any requirements for destroying data at end of study:
None provided

Type of data provided:
(e.g. self-report survey / medical record (MRR) / qualitative / self-report trial / objective trial data / 'other' - please describe)

Anonymous dataset of 7 variables at all follow up time points

Does the data set contain any personally identifiable data?	**YES**		**NO**	**x**
If yes, please specify:				

Expected server path / file location to hold data at Keele:

Pre87 S:\CHp 5 exploratory analysis

Keele iPCHS / CTU Lead Researcher: please complete the form and email a copy of completed form and provide signed a hard copy to the Keele CTU Data Systems Manager).

Please note on receiving this data into the iPCHS / Keele CTU you (Keele PI) are agreeing to:

- adhere to the Research Institute for Primary Care and Health Sciences (iPCHS) and CTU (where appropriate) data management policy, data security and applicable procedures.
- take full responsibility for the data held in the centre.
- comply with the principles and conditions set out in the Data Protection Act 1998, the Research Governance Framework and maintain proper safeguards to ensure confidentiality.
- only use the data for the approved purpose of the release and not allow further data release
- satisfy acknowledgement, co authorship and publication permissions (if applicable)

Name of Keele PI	**Signature**	**Date**
L Wood	*L Wood*	**5/11/2018**

Research Institute for Primary Care and Health Sciences (iPCHS) and Keele CTU, Keele University

Record of external data stored at the iPCHS / Keele CTU provided by an external source (e.g. another university or NHS organisation)

To be completed by the Keele Researcher of the proposed study

The lead iPCHS / CTU researcher should complete this form in conjunction with the Principal Investigator (PI) of the collaborating unit which is providing the external data to be stored at Keele iPCHS / CTU

Proposed Study Title: *for which data is being collected*	Exercise, as a complex intervention for chronic low back pain – does matching the primary outcomes to treatment targets matter?	
Collaborative Research? *Pease tick ✓ selection*		**No** collaboration, data/data provision to Keele with no collaboration

Keele Study Team:

This section of the form will include members of the iPCHS / CTU and will be used to support decisions on authorship of papers generated from this secondary analysis (if applicable)

Details	*Name*	*Role in study team*	*Contact details*
	L Wood	PhD Student	L.wood2@keele.ac.uk
	A Bishop	Lead Supervisor	a.bishop@keele.ac.uk
	N Foster	Supervisor	n.foster@keele.ac.uk
	M Lewis	Supervisor	a.m.lewis@keele.ac.uk

Institution/Organisation providing data:

Please provide details of the data owner/custodian and personnel providing approval of data release.

External Research Team Details	*Name*	*Role in study team*	*Contact details*
Data custodian Brazil, Sao Paulo	Gisela Miyamoto	Lead Author	gfisio_miyamoto@hotmail.com

Details of study providing dataset:

Study name	Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation
Sponsor	Universidade Cidade de Sao Paulo, São Paulo, Brazil
Funder	GCM was granted a PhD scholarship from São Paulo Research Foundation (FAPESP)
Funder Reference No.	**2013/26321-8 and 2016/07915-2**

Proposed Study Details

Research Question / Objective: Does a composite primary outcome matched to the treatment targets of the intervention change the conclusion(s) of the RCT?	
Analysis planned Start Date: October 2018	**End date: June 2019**

In to which iPCHS programme(s) does this study fit? *(Please tick all that apply)*

Osteoporosis and Osteoarthritis		**Prognosis and Consultation Epidemiology**	
Musculoskeletal Health and Pain	x	**Education & Training**	
Inflammatory studies		**Secondary Care**	
Mental Health studies		**Other (please describe)**	

Brief summary of process gone through to gain approval to hold data:
Attach copies of supporting documents submitted to external institution and confirmation of approval (e.g. data sharing agreement.)
Please see attached protocol

Are there any specific conditions applied to the use of the data by the collaborating site:	**YES**		**NO**	**x**
If yes, please specify:				

Details of any requirements for destroying data at end of study:
None provided

Type of data provided:
Anonymous dataset of 7 variables at all follow up time points

Does the data set contain any personally identifiable data?	**YES**		**NO**	**x**
If yes, please specify:				

Expected server path / file location to hold data at Keele:

Pre87 S:\CHp 5 exploratory analysis

Keele iPCHS / CTU Lead Researcher: please complete the form and email a copy of completed form and provide signed a hard copy to the Keele CTU Data Systems Manager).

Please note on receiving this data into the iPCHS / Keele CTU you (Keele PI) are agreeing to:

- adhere to the Research Institute for Primary Care and Health Sciences (iPCHS) and CTU (where appropriate) data management policy, data security and applicable procedures.
- take full responsibility for the data held in the centre.
- comply with the principles and conditions set out in the Data Protection Act 1998, the Research Governance Framework and maintain proper safeguards to ensure confidentiality.
- only use the data for the approved purpose of the release and not allow further data release
- satisfy acknowledgement, co authorship and publication permissions (if applicable)

Name of Keele PI	**Signature**	**Date**
L Wood	*L Wood*	**9/10/2018**

i. Linear mixed model results of Miyamoto et al. (2018) trial dataset: all time-points

	Pilates 1 vs CG			**Pilates 2 vs CG**			**Pilates 3 vs CG**			**Intervention (ALL) vs CG**	
	Effect Estimate	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig.**	**t-score**	**Effect Estimate**	**Sig**
Standardised Pain (PRIMARY)*											
6-week follow-up	0.43 (0.04, 0.82)	0.029φ	2.193	0.88 (0.50, 1.27)	<0.0001	4.503	0.74 (0.35, 1.13)	<0.001	3.736	0.69 (0.37, 1.01)	<0.0001
6-month follow-up	0.09 (-0.30, 0.48)	0.652	0.452	0.39 (0.00, 0.77)	0.051	1.962	0.21 (-0.18, 0.60)	0.284	1.073	0.23 (-0.09, 0.55)	0.160
12-month follow-up	0.04 (-0.36, 0.44)	0.832	-0.212	0.29 (-0.11, 0.69)	0.161	1.407	0.08 (-0.32, 0.48)	0.698	0.389	0.11 (-0.22, 0.44)	0.524
Standardised Composite Outcome *											
6-week follow-up	0.44 (0.21, 0.66)	**<0.001**	**3.806**	0.77 (1.00, 0.55)	<**0.0001**	**6.811**	0.59 (0.36, 0.81)	**<0.0001**	**5.079**	0.59 (0.41, 0.78)	<0.0001
6-month follow-up	0.04 (-0.18, 0.28)	0.704	0.381	0.34 (0.58, 0.11)	**0.004**	**2.911**	0.23 (0.00, 0.47)	0.051	**1.957**	0.21 (0.01, 0.40)	**0.037**
12-month follow-up	**0.09** (-0.34, 0.16)	**0.468**	**0.726**	0.24 (0.01, 0.49)	0.057	1.911	0.02 (-0.23, 0.27)	0.881	0.150	0.06 (-0.15, 0.26)	0.597
Individual target-related outcomes included within the composite outcome:											
Standardised Physical Function*											
6-week follow-up	0.29 (-0.04, 0.61)	0.081	1.749	0.82 (0.50,1.14)	p<0.0001	5.032	0.53 (0.20,0.85)	0.001	3.206	0.55 (0.27, 0.81)	<0.001
6-month follow-up	-0.04 (-0.36, 0.28)	0.794	-0.261	0.39 (0.07, 0.72)	0.017	2.406	0.22 (-0.10, 0.54)	0.182	1.339	0.19 (-0.08, 0.46)	0.167
12-month follow-up	0.08 (-0.42, 0.25)	0.625	-0.489	0.3 (-0.04,0.63)	0.085	1.729	0.05 (-0.29, 0.39)	0.78	0.279	0.09 (-0.19, 0.37)	0.543
Standardised Catastrophising*											
6-week follow-up	0.23 (-0.09, 0.54)	0.155	1.426	0.49 (-0.18, 0.81)	0.002	3.127	0.18 (-0.13, 0.50)	0.261	1.127	0.30 (0.05, 0.56)	0.021
6-month follow-up	0.10 (-0.24, 0.43)	0.566	0.575	0.38 (0.04, 0.71)	0.027	2.224	0.23 (-0.11, 0.56)	0.180	1.345	0.24 (-0.04,0.51)	0.094
12-month follow-up	-0.22 (-0.56, 0.12)	0.198	-1.291	0.34 (0.01, 0.67)	0.046	2.006	0.10 (-0.23, 0.44)	0.550	0.599	0.08 (-0.20, 0.36)	0.571

Standardised Kinesiophobia*											
6-week follow-up	0.40 (0.09, 0.72)	0.012	2.524	0.50 (0.19, 0.81)	0.002	3.159	0.44 (0.12, 0.75)	0.007	2.727	0.45 (0.19, 0.70)	0.001
6-month follow-up	0.12 (-0.22, 0.46)	0.496	0.682	0.29 (-0.05, 0.63)	0.094	1.683	0.04 (-0.30, 0.39)	0.797	0.258	0.15 (-0.13,0.43)	0.289
12-month follow-up	-0.08 (-0.47, 0.30)	0.672	-0.423	0.20 (-0.18, 0.58)	0.299	1.04	-0.33(-0.71, 0.05)	0.088	1.713	-0.07 (-0.39, 0.25)	0.665
Standardised GPE											
6-week follow-up	0.77 (0.35, 1.19)	<0.001	3.615	1.23 (0.81, 1.64)	<0.00001	5.769	1.11(0.69, 1.54)	<0.00001	5.184	1.04 (0.69, 1.38)	<0.0001
6-month follow-up	0.20 (-0.22, 0.63)	0.347	0.943	0.56 (0.13,0.99)	0.011	2.569	0.66 (0.23, 1.09)	0.003	3.015	0.47 (0.12,0.83)	0.009
12-month follow-up	-0.04 (-0.46, 0.38)	0.864	0.171	0.30 (-0.12, 0.72)	0.157	1.421	0.37 (-0.06,0.79)	0.09	-1.704	0.21 (-0.14, 0.56)	0.238
Standardised PSFS											
6-week follow-up	0.51(0.12, 0.91)	0.01	2.58	0.72 (0.33, 1.11)	0.0003	3.651	0.54 (0.14, 0.93)	0.01	2.58	0.59 (0.27, 0.91)	0.0003
6-month follow-up	-0.28 (-0.68, 0.13)	0.179	-1.348	0.12 (-0.29,0.52)	0.572	0.566	0.04 (-0.37, 0.45)	0.841	0.2	-0.04 (-0.38, 0.29)	0.793
12-month follow-up	-0.09 (-0.54, 0.36)	0.706	-0.378	0.21 (-0.23, 0.66)	0.602	0.956	-0.12 (-0.56, 0.33)	0.602	-0.523	0.00 (-0.36, 0.37)	0.980

Bold items indicate where the composite value t-score is greater than that of the original primary outcome with a smaller p-value. CG is control group; Pilates 1 is Pilates once a week, Pilates 2 is Pilates twice weekly, Pilates 3 is Pilates thrice weekly, Intervention ALL is the combined effect of all three intervention arms, CI is confidence interval. GPE is global perceived effect, PSFS is Patient-Specific Functional Scale; All values are calculated as the mean (intervention) minus the mean (control) values, where positive values favour the intervention arm. Outcomes with an * have scales running in opposite direction to others and have been multiplied by minus one to trend in a positive direction. φ This P-value does not match exactly with the 0.007 in Table 3 possibly due to the slightly different correlation structures that are modelled in (perhaps these may also have been slightly different in specification i.e. unstructured vs autoregressive)

Table to show the standard error of the two outcome variables

Parameter	Standardised Pain Outcome	Standardised Composite Outcome
FU1 Pilates 1 vs CG	0.08	0.04
FU1 Pilates 2 vs CG	0.06	0.03
FU1 Pilates 3 vs CG	0.09	0.05
FU2 Pilates 1 vs CG	0.06	0.04
FU2 Pilates 2 vs CG	0.07	0.04
FU2 Pilates 3 vs CG	0.09	0.05
FU3 Pilates 1 vs CG	0.06	0.03
FU3 Pilates 2 vs CG	0.06	0.03
FU3 Pilates 3 vs CG	0.08	0.03

CG is control group, FU is the follow up time-point where FU1 is 6-weeks, FU2 is 6-months and FU3 is 12-months

j. Linear mixed model results of Moffett et al. (2006) trial dataset: all time-points

	Adjusted mean difference (95% CI)		
	McKenzie vs SFA		
Standardised Fear-Avoidance (TSK-AA) *	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	-0.01 (-0.22,0.20)	0.94	-0.079
6-month follow-up	***0.25 (0.05,0.45)***	***0.01***	***2.475***
12-month follow-up	0.09 (-0.30, 0.12)	0.39	0.854
Standardised Composite Outcome	**Effect Estimate**	**Sig.**	**t-score**
6-week follow-up	-0.01 (-0.11,0.09)	0.868	-0.167
6-month follow-up	0.08 (-0.02,0.19)	0.124	1.544
12-month follow-up	0.01 (-0.11, 0.12)	0.922	0.097
Individual Components of the Composite			
Standardised fear-avoidance (TSK-SF and TSK-AA combined) *			
6-week follow-up	-0.11 (-0.28, 0.06)	0.214	-1.25
6-month follow-up	0.20 (0.02, 0.37)	0.028	2.21
12-month follow-up	0.08 (-0.1, 0.26)	0.388	0.87
Standardised physical function *			
6-week follow-up	0.14 (-0.09,0.37)	0.227	1.21
6-month follow-up	0.10 (-0.11, 0.31)	0.356	0.93
12-month follow-up	-0.16 (-0.41, 0.09)	0.214	-1.25
Standardised health control (internal)			
6-week follow-up	-0.03 (-0.16, 0.22)	0.763	0.302
6-month follow-up	0.12 (-0.07, 0.31)	0.225	-1.215
12-month follow-up	-0.04 (-0.25, 0.18)	0.734	-0.340
Standardised health control (chance scale) *			
6-week follow-up	0.08 (-0.11,0.27)	0.413	-0.820
6-month follow-up	0.08 (-0.14, 0.30)	0.469	-0.726
12-month follow-up	0.00 (-0.20, 0.21)	0.984	-0.021
Standardised health control (powerful others scale) *			
6-week follow-up	-0.07 (-0.25,0.12)	0.491	-0.690
6-month follow-up	0.04 (-0.19, 0.27)	0.743	0.328
12-month follow-up	0.24 (0.00, 0.47)	0.049	1.977
Standardised self-efficacy			
6-week follow-up	0.01 (-0.20, 0.18)	0.925	0.094
6-month follow-up	0.14 (-0.34, 0.06)	0.167	1.385
12-month follow-up	-0.09 (-0.12, 0.30)	0.395	-0.852
Standardised anxiety *			
6-week follow-up	0.08 (-0.09, 0.24)	0.357	0.92
6-month follow-up	-0.05 (-0.24, 0.14)	0.605	-0.519
12-month follow-up	-0.11 (-0.3, 0.08)	0.269	-1.107
Standardised depression*			
6-week follow-up	0.05 (-0.13, 0.22)	0.590	0.540
6-month follow-up	0.04 (-0.14, 0.22)	0.647	0.458
12-month follow-up	-0.04 (-0.23, 0.15)	0.678	-0.416

***Bold italics items indicate where statistical significance; Where SFA represents Solution Finding Approach and TSK represents the Tampa Scale of Kinesiophobia; TSK-AA is the Activity Avoidance subscale and the TSK-SF is the somatic focus subscale. Asterisks denote*

the scale has been multiplied by minus one to trend in a positive direction. Positive values favour the intervention arm.

Table to show the standard error of the two outcome variables

Parameter	Standardised Fear-Avoidance Outcome	Standardised Composite Outcome
1,1	0.06	0.02
2,1	0.06	0.02
2,2	0.07	0.02
3,1	0.05	0.02
3,2	0.06	0.02
3.3	0.07	0.02

k. Linear mixed model results of Groessl et al. (2017) trial dataset: all time-points

Table to show the linear mixed model analysis of the standardised pain outcome in comparison to the composite outcome results

	Adjusted mean difference (95% CI)		
Yoga vs WL	**Effect estimate (95% CI)**	**t score**	**Sig.**
Standardised Primary Outcome (Physical Function) (trial results)			
6-week follow-up	-0.13 (-0.45,0.19)		0.37
12-week follow-up	-0.14 (-0.46,0.18)		0.34
6-month follow-up	*-0.45 (-0.77, -0.13)*		*0.003*
Standardised Composite Analysis			
6-week follow-up	*-0.13 (-0.31, 0.05)*	*-1.431*	*0.155*
12-week follow-up	*-0.25 (-0.43, -0.07)*	*-2.761*	*0.007*
6-month follow-up	*-0.30 (-0.55, -0.04)*	*-2.324*	*0.022*
Individual Components of the Composite Outcome:			
Standardised Pain			
6-week follow-up	*-0.35 (-0.57, -0.14)*	*-3.3*	*0.001*
12-week follow-up	*-0.30 (-0.52, -0.08)*	*-2.741*	*0.007*
6-month follow-up	*-0.27 (-0.5, -0.03)*	*-2.217*	*0.029*
Standardised Plank			
6-week follow-up	-0.00 (-0.33, 0.32)	-0.025	0.980
12-week follow-up	-0.23 (-0.51, 0.04)	-1.64	0.105
6-month follow-up	-0.33 (-0.68, 0.00)	-1.929	0.057
Standardised Flexion ROM			
6-week follow-up	0.14 (-0.23, 0.51)	0.746	0.457
12-week follow-up	-0.27 (-0.61, 0.08)	-1.538	0.127
6-month follow-up	-0.03 (-0.41, 0.36)	-0.151	0.88
Standardised Extension ROM			
6-week follow-up	-0.29 (-0.61, 0.02)	-1.86	0.065
12-week follow-up	-0.08 (-0.44, 0.28)	-0.456	0.649
6-month follow-up	-0.21 (-0.64, 0.22)	-0.951	0.344

**Where WL is waiting list, ROM is range of motion. All values are calculated as mean difference of yoga minus mean difference of waiting list, where negative values favour the yoga intervention. Shaded values reflect the primary time-point.*

Table to show the standard error of the two models

Parameter	Standardised Pain Outcome	Standardised Composite Outcome
1,1	0.12	0.06
2,1	0.11	0.05
2,2	0.12	0.05
3,1	0.11	0.04
3,2	0.16	0.04
3,3	0.14	0.05
4,1	0.10	0.04
4,2	0.10	0.04
4,3	0.11	0.04
4,4	0.11	0.04

I. Additional Information Regarding Consensus Workshops

Ethical approval for first workshop:

Keele UNIVERSITY

17/08/2018

Dear Lianne

PI: Lianne Wood
Title: The treatment targets and outcomes of exercise as an example of a complex intervention in chronic low back pain
Ref: ERP2393

Thank you for submitting your application for review. The proposal was reviewed by the Panel Chair. I am pleased to inform you that your application has been approved by the Ethics Review Panel.

If the fieldwork goes beyond the date stated in your application, or there are any amendments to your study you must submit an 'application to amend study' form to the ERP administrator at research.governance@keele.ac.uk. This form is available via https://www.keele.ac.uk/raise/researchsupport/projectassurance/researchethics/

If you have any queries please do not hesitate to contact me, in writing, via the ERP administrator, at research.governance@keele.ac.uk stating **ERP2393** in the subject line of the e-mail.

Yours sincerely
P.P.

K. Cunningham

Dr Colin Rigby
Chair – Ethical Review Panel

Amendment to ethical approval:

Keele University FMHS Faculty Research Ethics Committee
health.ethics@keele.ac.uk

15 April 2019

Dear Lianne,

Project Title:	The treatment targets and outcomes of exercise as an example of a complex intervention in chronic low back pain.
REC Project Reference:	MH-190021
Type of Application	Amendment to MH-190019

Keele University's Faculty of Medicine and Health Sciences Research Ethics Committee (FMHS FREC) reviewed the above amendment.

Favourable Ethical opinion
The members of the Committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation; there are no conditions attached to this opinion.

Reporting requirements
The University's standard operating procedures give detailed guidance on reporting requirements for studies with a favourable opinion including:

- Notifying substantial amendments
- Notifying issues which may have an impact upon ethical opinion of the study
- Progress reports
- Notifying the end of the study

Approved documents
The documents reviewed and approved are:

Document	Version	Date
Those submitted with Amendment application MH-190021		

Yours sincerely,

Ed Chadwick

Dr Ed Chadwick
Committee Chair

Informed consent form for national workshop:

Consensus Workshop, October 2018

Keele UNIVERSITY

CONSENT FORM

Please initial all the boxes if you agree

1. I confirm that I have read and understand the study information leaflet (dated 8/18) and am willing and able to take part in the study ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. I understand that if I do withdraw, my contribution will still remain valid. ☐

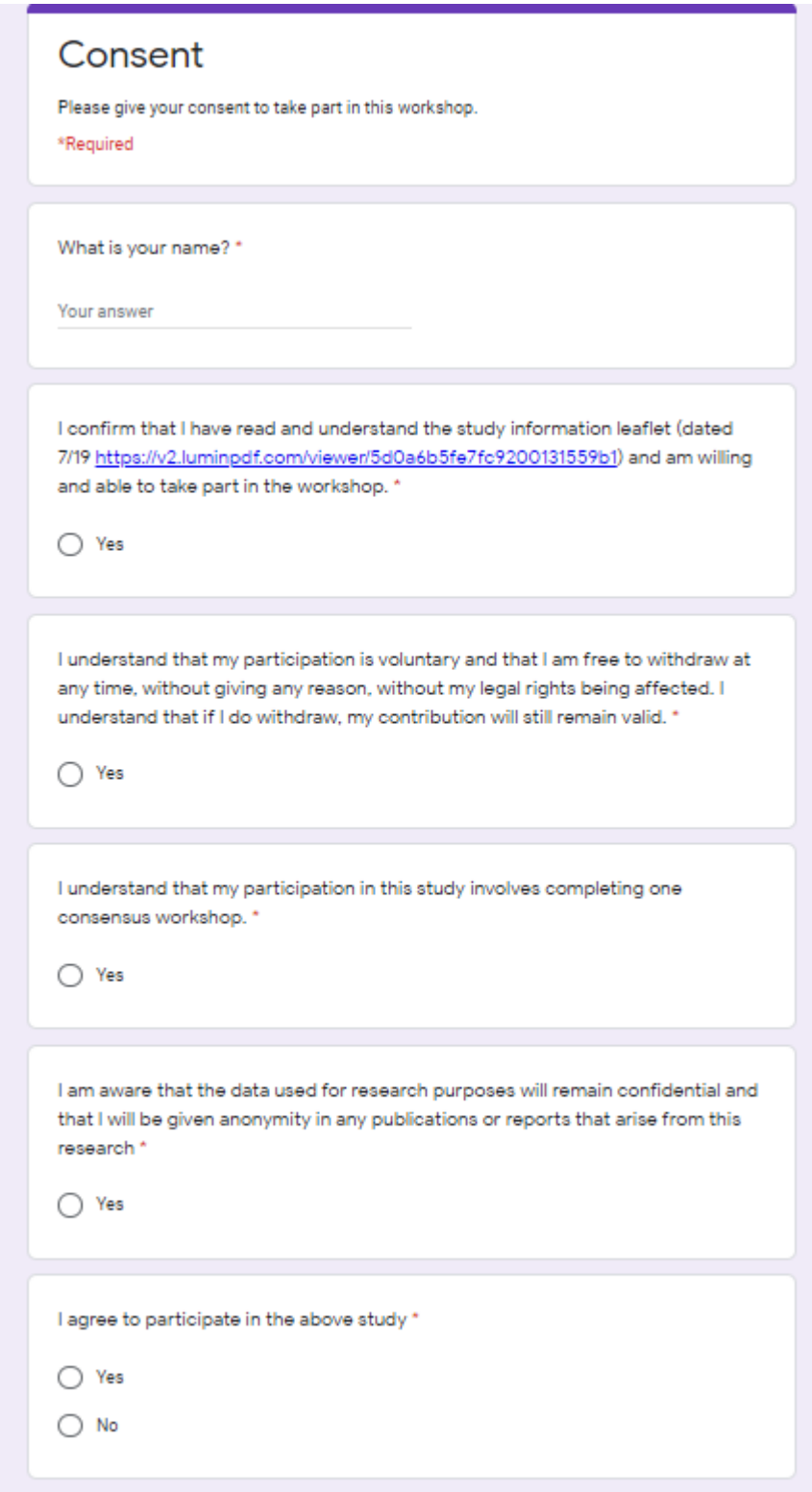
3. I understand that my participation in this study involves completing one consensus workshop ☐

4. I am aware that the data used for research purposes will remain confidential and that I will be given anonymity in any publications or reports that arise from this research ☐

5. I agree to participate in the above study ☐

Name of participant (print)	Date	Signature
.....	.....	.....
Name of researcher (print)	Date	Signature
.....	.....	.....

Informed consent form for international workshop:



Consent

Please give your consent to take part in this workshop.

*Required

What is your name? *

Your answer

I confirm that I have read and understand the study information leaflet (dated 7/19 https://v2.luminpdf.com/viewer/5d0a6b5fe7fc9200131559b1) and am willing and able to take part in the workshop. *

- Yes

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. I understand that if I do withdraw, my contribution will still remain valid. *

- Yes

I understand that my participation in this study involves completing one consensus workshop. *

- Yes

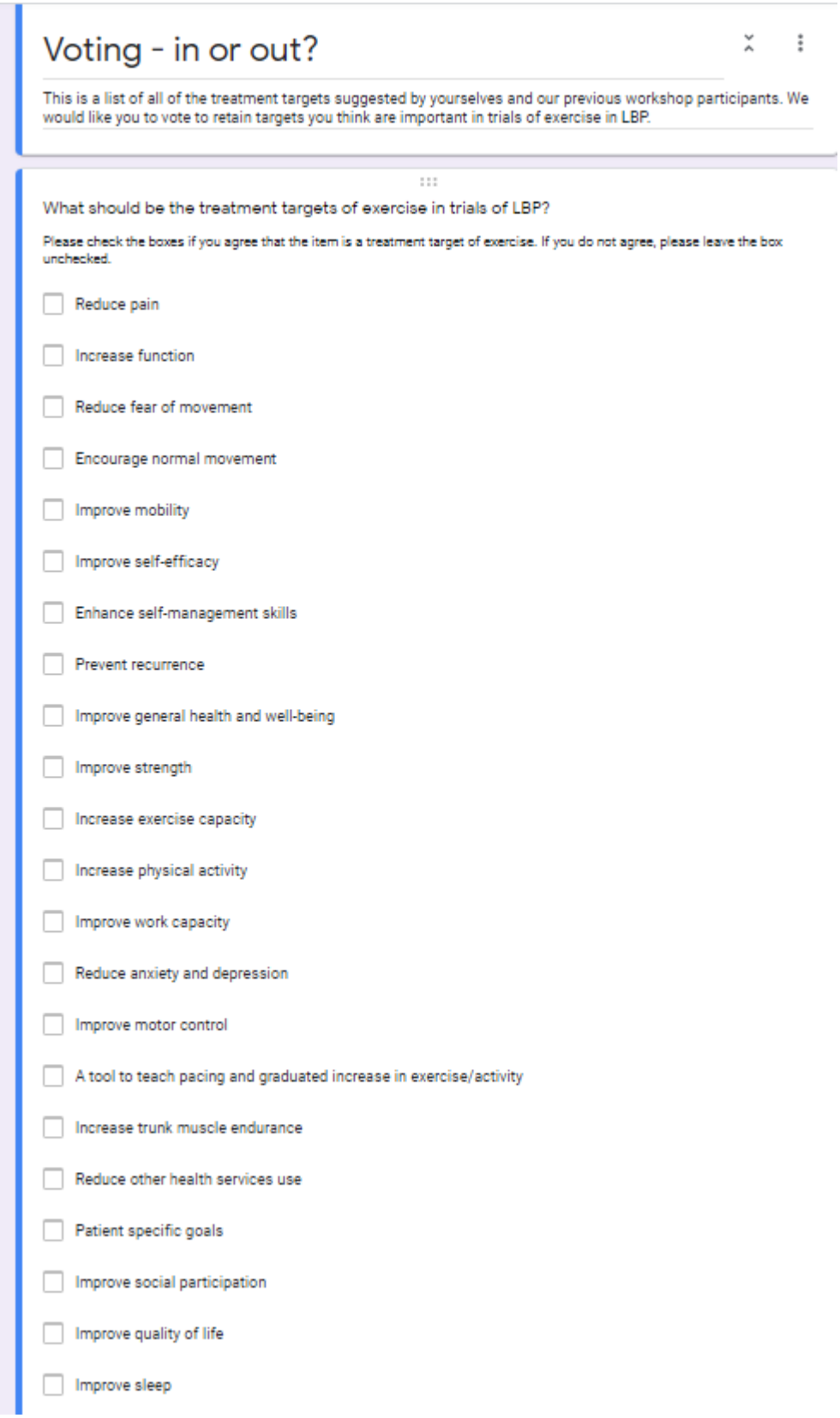
I am aware that the data used for research purposes will remain confidential and that I will be given anonymity in any publications or reports that arise from this research *

- Yes

I agree to participate in the above study *

- Yes
- No

Google sheets voting form used in International Workshop:



Voting - in or out?

This is a list of all of the treatment targets suggested by yourselves and our previous workshop participants. We would like you to vote to retain targets you think are important in trials of exercise in LBP.

What should be the treatment targets of exercise in trials of LBP?

Please check the boxes if you agree that the item is a treatment target of exercise. If you do not agree, please leave the box unchecked.

- [ ] Reduce pain
- [ ] Increase function
- [ ] Reduce fear of movement
- [ ] Encourage normal movement
- [ ] Improve mobility
- [ ] Improve self-efficacy
- [ ] Enhance self-management skills
- [ ] Prevent recurrence
- [ ] Improve general health and well-being
- [ ] Improve strength
- [ ] Increase exercise capacity
- [ ] Increase physical activity
- [ ] Improve work capacity
- [ ] Reduce anxiety and depression
- [ ] Improve motor control
- [ ] A tool to teach pacing and graduated increase in exercise/activity
- [ ] Increase trunk muscle endurance
- [ ] Reduce other health services use
- [ ] Patient specific goals
- [ ] Improve social participation
- [ ] Improve quality of life
- [ ] Improve sleep

List of the treatment targets after idea generation stage in National Workshop

Treatment Targets from SR (n=30)	**Added Brainstormed Treatment Targets (n=25)**
Reduce back pain	Reduce Pain
Spinal stabilisation / Spinal control altered /Trunk stability	A tool to teach pacing and graduated increase in exercise/ activity
Strengthening/ Muscle strength	Decrease barriers to movement
Stretching/ Flexibility	Decrease threat
Posture	Enhance self-management skills
Self-confidence	Ensure mobility into the future
Functional capacity/ improve function	Improve fitness
Improve physical capability / activity	Improve mobility of the spine
Increase trunk muscle endurance	Improve motor control
Disability	Improve physical activity levels
Reduce stress	Improve proprioception
Recovery	Improve strength
Fear of movement associated with pain	Improve work capacity
Reduce deconditioning	Increase exercise compliance
Prevent recurrence and chronicity	Increase function
Balance	Increase physical activity and or exercise capacity
Body awareness	Increase range of movement - spinal and other joints
General health and well-being	Limit time to return to work (full, partial duties)
Muscle tension	Optimise neural function
Relaxation	Reduce dependence on health service
Encourage normal movement	Reduce need for surgery
Self-efficacy; improved control	Reduce other health services use (other treatments/ medication/ testing)
Psychosocial factors	Reduce pain
Mobility	Restore neural mobility
Mental positivity	Stretch the connective tissue
Reduce anxiety and depression	
Reduce absence from work	
Catastrophising	
Kinesiophobia	
Improved coping	

Table to Demonstrate Grouping of Treatment Targets in National Workshop

Treatment Target	**Grouped Targets**
Reduce pain	Reduce back pain
Reduce fear of movement	Decrease threat
	Reduce kinesiophobia
Increase function and reduce disability	Improve functional capacity
	Reduce disability
Improve self-efficacy	Increase self-esteem
	Improve self-confidence
	Improve coping ability
Increase physical activity and/ or exercise capacity	Improve fitness
	Improve physical capability
	Improve physical activity levels
	Reduce deconditioning
Improve recovery	Prevent recurrence
Improve mobility	Increase range of movement - spinal and other joints
	Help ensure mobility into the future
	Improve mobility spine
	Muscle flexibility
	Stretch the connective tissue
	Restore neural mobility
	Optimize neural function
Improve strength	Muscle strengthening
Reduce other health services use (medications/ testing/ treatments)	Reduce need for surgery
	Reduce dependence on health service
Increase work capacity and ADLs	Reduce absence from work
	Limit time to return to work

Treatment targets generated in idea generation stage of international workshop

	Treatment Target	Grouped Targets within these Treatment Targets
Targets generated by final ranking of national workshop	Reduce pain	Reduce back pain
	Increase function	Improve functional capacity
		Reduce disability
	Reduce fear of movement	Decrease barriers to movement
		Decrease threat
		Reduce kinesiophobia
	Encourage normal movement	
	Improve mobility	Increase range of motion- spinal and other joints
		Help ensure mobility into the future
		Improve mobility of the spine
		Muscle flexibility
		Stretch the connective tissue
		Restore neural mobility
		Optimise neural function
	Improve self-efficacy	Improve self-confidence
		Increase self-esteem
		Improve coping ability
	Enhance self-management skills	
	Prevent recurrence	Improve recovery
	Improve general health and well-being	
	Improve strength	Improve muscle strength
	Increase exercise capacity	Reduce deconditioning
		Improve fitness
	Increase physical activity	Improve physical capability
		Improve physical activity levels
	Improve work capacity	Reduce absence from work
		Limit time to return to work (full, partial duties)
	Reduce anxiety and depression	Improve mental positivity
	Improve motor control	
	A tool to teach pacing and graduated increase in exercise/activity	
	Increase trunk muscle endurance	
	Reduce other health services use	Reduce need for surgery
		Reduce dependence on health service
Targets added in idea generation	Patient-specific goals	
	Improve social participation	
	Improve quality of life	
	Improve sleep	
	Improve cognitive function	
	Lower inflammation	
	Behavioural change	
	Weight loss/ gain	
	Improve education and/or knowledge	

	Improve functional mobility
	Improve performance
	Improve attitudes/ cognitions/ beliefs
	Reduce catastrophising
	Increase body awareness
	Increase body image

m. Summary of Protocols of 23 Current RCTs

Registry Updated	Trial	Intervention	Treatment Target	Outcome (Primary outcome in bold)	Categorisation
June 2020	ISRCTN55789697	Task-orientated exercises and CBT concerning beliefs compared to exercises for passive spinal mobilisation, strengthening, stretching and postural control	Modifying wrong beliefs and behaviours concerning subacute pain	**Physical function (ODI);** Pain intensity (NRS), Kinesiophobia (TSK), unhelpful beliefs (Pain Beliefs and Perceptions Inventory), Anxiety and depression HADS), strategies for coping pain (CSQ-R); HRQoL (SF-36), RTW, Sick leave days, Efficacy of treatment (GPE)	Unmatched
May 2019	CT03200509	Supervised exercise and health coaching with activity monitor	To identify facilitators and barriers to physical activity participation and to assist participants to achieve their physical activity goals by providing ongoing education and support	**Physical activity (counts per min, disability, pain intensity);** objective measures physical activity (step counts, time spent in moderate/ light/ vigorous activity); Self-reported physical activity levels; Self-reported sedentary behaviour; Depression, HRQoL, PSEQ Weight related outcomes	Matched
May 2019	ISRCTN12965286	CFT feasibility study	"individualised self-management intervention targets psychological, physical and lifestyle barriers"	**Physical Function (RMDQ);** pain intensity (NRS); fear-avoidance (FABQ); Keele STarT Back Screening Tool; self-efficacy (PSEQ); pain catastrophising (PCS); Distress, Anxiety and Stress Scale (DASS 21); HRQoL (Eq-5D-5L); Global Rating of Change (GPE); participant satisfaction; Working Alliance Theory of Change Inventory	Unmatched
June 2020	ISRCTN17816427	ALL MSK PAIN	to increase physical activity	**Average daily step count, Pain intensity and location (NRS, pain**	Matched

		Usual care, pedometer care, Ipopp walking intervention		**mannequin);** Physical function (PF-10 from the SF-36); Physical activity (IPAQ-E); Sedentary time, (accelerometry); HRQoL (EQ-5D-5L); Non-health aspects of quality of life (ICECAP-O); Self-efficacy (Self-Efficacy for Exercise Scale); Anxiety (GAD-7); Depression (PHQ 8); Health-care resource use (patient self -reported questionnaires); Adverse events; Pedometer use (patient self -reported questionnaires); Consultation experience (iPOPP intervention only)	
March 2019	ISRCTN14736486	Usual care 2. Usual care + internet intervention 3. Usual care + internet intervention + telephone physiotherapist support	? encourage self-management, increase physical activity (inferred)	**Physical function (RMDQ);** health economics (HRQoL, over the counter medication use, participant borne costs, occupational items, health-care resource use) Pain (number troublesome days, NPRS, risk persistent disability Start Back score), Psychological outcomes – fear of movement, negative orientation (PCS), confidence in managing (PSEQ) MSK HQ, PHQ mental health Physical activity: Gdin Leisure- Time exercise questionnaire	Unmatched
April 2020	ISRCTN94074203	Physical therapy with and without an app.	Improve self-management, promote recovery, resume daily activities. Increase adherence to physical activity.	**Physical function (ODI), long term reduction low back pain related costs using cost questionnaires – health-care utilisation, productivity losses.** Pain intensity, physical activity, adherence, psychological function (PCS, FABQ), self-efficacy, self-management skills, HRQoL, Start Back Tool, Central sensitivity,	Unmatched
Jan 2020	ISRCTN15830360	Chiropractic, PT, combined treatment and UC	None specified	**Physical Function (ODI),** Pain intensity (NRS), General health, (self-rated health questionnaire), HRQoL (EQ-5D), working status (% Full-time work)	Unmatched

Nov 2019	ISRCTN42338218	12 week digital care pathway vs UC	None specified	**Pain and disability using modified von Korff scales and ODI**; LBP (VAS), Impact of pain on daily life (CAS), surgery interest and intent, ability to self-manage	
Dec 2018	ISRCTN14136384	MDT care, Physio	Aiming at full or partial RTW	**RTW (first 4-week period)**,physical function (RMDQ)	Matched
May 2019	ISRCTN99926592	Green exercise, balneotherapy, combination, or CG	Unclear ? reduce symptoms and improve wellbeing	**Functional spine mobility (MediMouse)**, Pain (mVAS), status of health (mVAS), Back performance scale, torso/spine rotation, HRQoL (SF-36), depression (WHO-5), claim medical care (medication, no of medical consultations), pain behaviour and activities of daily living (ODI)	Unmatched
Jan 2020	ISRCTN11875357	Start back stratified care.	Unclear	**Physical Function (ODI at 12-months),** pain and disability (ODI at 3-months, RMDQ, PROMIS, frequency LBP during past 3 months, LBP intensity (NRS), leg pain intensity (NRS), STarT back, HRQoL(EQ-5D), Direct costs and indirect costs.	Unmatched
Oct 2019	NCT03424278	Motor control vs resistance training	MCE “evidence motor control dysfunction” Strength building exercise	**Pain (NRS), Physical function (RMDQ), Kinesiophobia (TSK),** Trunk strength	Unmatched
Mar 2020	NCT03753165	High intensity exercise + PT, PT only	Inferred “improving the physical fitness, peak performance, autonomic balance, muscle strength and coordination in athlete”	**Change in heart rate variability (heart rate variability parameters)**,Heart rate recovery (post-exercise), Pain intensity(NRS), physical function (ODI), Arterial Blood pressure (Baroreceptor sensitivity)	Unmatched
Sept 2019	NCT03113292	Pilates vs HEP	Unclear	**Pain intensity (VAS) physical function (QBPDS)**, Health status (EQ-5D), perception of recovery (GPE), postural balance (Balance platform)	Unmatched

Feb 2020	NCT04000685	Yoga vs stabilisation exercise vs aerobic walking.	None	**Pain severity (VAS), physical functional (ODI),** HRQoL (Nottingham Health Profile), gait parameters (Gait assessment), metabolic capacity (modified Bruce protocol), cognitive level (Montreal Cognitive Assessment Questionnaire), alexithymia (Toronto Alexithymia scale), kinesiophobia (FABQ), back awareness. (Freemantle Back Awareness Questionnaire)	Unmatched
2016	NCT02703402	Ex vs HEP	None	**Pain (VAS),** HRQoL (SF-36), functional capacity (six minute walk test), flexibility (sit and reach test)	Unmatched
Feb 2019	NCT03778970	Ex + pain neuroscience education vs Ex + Education	"posture and control of movements which are impaired and trigger pain."	**Pain intensity (NRS), disability (ODI),** pain catastrophising (PCS), pain self-efficacy (PSEQ), fear-avoidance beliefs (FABQ), Exercise adherence, global perceived effect (GPE)	Unmatched
Aug 2014	NCT02222935	Balance ex vs routine back ex programme	Balance deficits common -	**Balance (Starr Edward Balance test)** Berg Balance score, pain intensity (VAS)	Matched
Dec 2017	NCT03376724	Functional exercise vs Back school	None	**Pain Intensity (NRS),** physical function (ODI, RMDQ); fear of movement (FABQ), HRQoL (SF-36), functional capacity (6 min walk test, TUG), Recovery (Likert scale), analgesia	Unmatched
Apr 2020	NCT02969785	Lumbar stabilisation vs strengthening	postural control and trunk neuromuscular activity	**Electromyography measurement (EMG estimates, Root mean square); b**alance (Force platform measurement); pain intensity (McGill Short version); physical function (ODI), fear-avoidance (Waddel questionnaire)	Unmatched
Jan 2020	NCT02895828	CSE vs General strengthening ex	None	**HRQoL (SF-36);** physical function (ODI); Trunk muscle activity (Surface EMG), lumbar segmental motion (radiography), pain intensity (NRS)	Unmatched
Apr 2019	NCT03324659	Meditation + ex, meditation	None	**Physical Function: RMDQ** Cutaneous sensation, fear-avoidance	Unmatched

				(FABQ), mindfulness (Freiberg mindfulness inventory), pain intensity (VAS), anxiety inventory (State-trait anxiety inventory), change in pressure sensation, heat unpleasantness	
Feb 2020	NCT04283409	Graded activity vs motor control exercise	MCE: retrain optimal control and coordination of the spine. Graded activity address modifiable contextual factors associated with pain experience (psychological factors). Primary goal to increase activity tolerance.	**Physical Function: ODI;** Physical Function (PSFS); Pain (NRS); HRQoL (EQ-5D-5L); Impact of Low back pain: PROMIS-9; Lumbar spine instability questionnaire, OREBRO LBP screening questionnaire, TSK, CSQ, Pain Detect Questionnaire, SMART clinical checklist, PPT Assessment, Publicly funded health-care costs, patient direct health-care costs, societal health-care costs.	Unmatched

n. Summary of Mediation Analyses in LBP

Summary of mediation analyses performed on trials of exercise interventions for back pain (adapted from Lee et al., 2017)

Study	**Intervention (exposure)**	**Mediators tested**	**Outcome**
Focht, Rejeski, Ambrosius, Katula, & Messier (2005)	Exercise (n=80) Dietary weight loss (n=82) Combination therapy (n76)	**Stair climbing self-efficacy** Walking self-efficacy	Mobility (walking and stair climbing) Pain
Leeuw et al. (2008)	Graded exposure in vivo (n=42) Graded activity (n=43)	**Pain catastrophising** **Perceived harmfulness of activities**	Functional disability Main complaints
Seymour, Hughes, Campbell, Huber, & Desai (2009)	Exercise therapy delivered by physiotherapists Exercise therapy delivered by exercise instructors	**Exercise adherence self-efficacy** (barriers adherence, time adherence) Self-management self-efficacy (exercise, pain management, symptom management) Attendance, exercise maintenance	Function (muscle strength, exercise capacity, physical function) Pain
Smeets et al. (2006)	CBT (n=55) Active physical therapy (n=52) Combination therapy (n=55)	**Pain catastrophising,** internal pain control	Disability, Pain, Depression, Patient-specific complaints
Hall et al. (2016)	Tai chi (n=51) WL (n=51)	**Pain catastrophising**	Pain intensity, pain bothersomeness, disability
Mansell et al. (2016)	Stratified care (n=93) Current best care (n=45)	**Distress** (as a latent variable characterised by fear-avoidance, pain catastrophising, anxiety and depression)	disability
Stevens et al. (2018)	Multi-faceted work place intervention	Fear Perceived muscle strength, Use of assistive devices **Perceived physical exertion**	Days with LBP LBP intensity Bothersome LBP days
Mansell, Hill, Main, Von Korff, & van der Windt (2017)	Back in Action	**Fear-avoidance beliefs**	Disability